

KW chromatography; dementia; memory; brain function.

XX OS Gallus domesticus.

XX PN JP08245413-A.

XX PD 24-SEP-1996.

XX PF 13-MAR-1995; 95JP-0081797.

XX PR 13-MAR-1995; 95JP-0081797.

XX PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

XX DR WPI; 1996-482127/48.

XX PT New neural spine stretching agents comprising heat shock protein 90
PT - useful as memory-improving agents, brain function activators or
PT for treating dementia

XX PS Example 1; Fig 2; 5pp; Japanese.

XX CC This is the amino acid sequence of the chicken heat shock protein-90
CC which can be used as a neural spine elongation factor. The protein was
CC isolated from chicken cells by conventional chromatography and separated
CC by SDS-PAGE. Peptide fragments were isolated from the gel and were used
CC to ascertain the amino acid sequence. The protein can be used as the
CC main ingredient in compositions e.g. to treat dementia, to improve memory
CC or to activate brain functions.

XX SQ Sequence 728 AA;

Query Match 4.7%; Score 99.5; DB 17; Length 728;
Best Local Similarity 21.3%; Pred. No. 0.58;
Matches 79; Conservative 57; Mismatches 144; Indels 91; Gaps 18;

QY 6 VPSSLSAELEAKLIRFGKK--KNTHSLFVFII--PENFKGCISGHGMDIAL---TEP 57
DB 390 lphlshremllqsklikvirknlvkkclelftelaeedkenykkfyegfknkklgheds 449
QY 58 LTMKMSNVVYKWTCPSTNVKTENATGPEELGL---PLQRSYSEHLGYFPTDL----- 108
DB 450 qnrkkisellryyt-----sasgdemslkdycctrmkenqkhvyyitgetkdqva 499
QY 109 -FACSESLRNGCLEL-----NASLSEFEKNKIS-----LHSSKEKLRE 149
DB 500 nsafverlrk-hgleviymiepidycvqqlkefgktlsvtkglelpedeeekkkqe 558
QY 150 RIKYCCQLRTLPLPYVGRKNDAAVLEATVDYVIREKI-SPAVMAQITEALQSNMRF 208
DB 559 ekkakffnlgkim-----kdilekvekvvvshrlvtspccivstvygwtanmer 608
QY 209 CKKQQTPIELSLPGTYMAQRENSVMSTYSPERGLOFLTNTCMNGCSTPDAESSLDEAVRV 268
DB 609 imkaq-----alrdnstmgymaakkhlel-----npd--hslietlq 644
QY 269 PS-SSASENAIGD---PYKTHISSAALSIN--SLHTVRYIS--KVTPSYDATAVTNQNI 319
DB 645 kaeadkndksvkdlvallyetalissgfslepbqthanriymikiglgldedddtaaeaa 704
QY 320 SIHLPSAMPPV 330
DB 705 spavteemppl 715

Search completed: January 19, 2001, 10:43:54
Job time: 33644 sec

PS Claim 1; Column 73-79; 53pp; English.

XX This is the amino acid sequence of the human origin of replication complex protein 1 (ORC1). The sequence was isolated using primers synthesised based on amino acid sequence homology between the

CC S. cerevisiae (W22224), K. lactis (W22230) and S. pombe (W22232)

CC proteins. The amplified fragment was then used to screen a phage

CC lambda cDNA library constructed from NTERD21, an embryonic carcinoma

CC cell line, to isolate the full length gene. The ORC proteins

CC (W22224-35) can be used to screen chemical libraries to identify lead

CC compounds useful in treatment and diagnosis of undesired cell growth,

CC e.g. cancer, infections, inflammation and hypersensitivity.

XX SQ Sequence 861 AA;

Query Match 4.8%; Score 100; DB 18; Length 861;

Best Local Similarity 21.3%; Pred. No. 0.69; Mismatches 139; Indels 100; Gaps 15;

Matches 79; Conservative 53;

Qy 55 TEPLTME-----KMSNV-----VKYWTTCPS--NTVKTENATGPEELGLPLQRSYSEHLGY 103

Db 226 thltparkrlelgnlgnpqmsqqtscasldspgrikrvafseitspskrsqdklqt 285

Qy 104 FPTDLFACSELRNGLELNASLSEFEKKNKISLLH-----SSKEKLRERIKY 153

Db 286 lspalkapektretg-----lsyteddkkaspehrilrtriaaasktidiree---- 333

Qy 154 CCEQLRTLPLPVKGRKNDAAVLEATVDYVKYIREKISPAVMAQITEALQSNMRECKKQ 213

Db 334 -----rtltplsgggr-----ssvpsvilkpenikkrdaekaq-neatstphri----- 379

Qy 214 TPIELSLPGTVMAQRENSVMSTYSPERGLOFTNTCWNGCSTPDAAESSLDEAVRPSSSA 273

Db 380 -----rrkssvltmnrirqlrflgns-----ksdqeekelipaeeisdss 421

Qy 274 SENAGDPYKTHISSAALSIN-----SLHTVRYKVTSPSYDATATVNQINISHLSP-S 325

Db 422 deeeastpplpraprtvsnrlsrslkslhtl-----tkvpkkslkprtprc 469

Qy 326 AMPVVSFSFLG-----TALLGWARRALHIPTVCNS-----FGRIKSTCLKFTLST 370

Db 470 aapqirsrlaapeasvleearlhsvsavgpeslpcrqeqfdqlynfveskllhtgtgc 529

Qy 371 TYWAQFDNLGK 381

Db 530 myisgvpgtgk 540

RESULT 14

W14138

XX W14138 standard; Protein; 861 AA.

AC W14138;

XX 23-JUL-1997 (first entry)

DE Human origin of replication complex ORC1.

XX Origin of replication complex; ORC; gene therapy; cancer;

KW neoplasia; inflammation; hypersensitivity.

XX Homo sapiens.

OS W09640977-A1.

XX W09640977-A1.

XX 19-DEC-1996.

PD 07-JUN-1996; 96WO-US09403.

XX 07-JUN-1995; 95US-0484105.

XX (COLD-) COLD SPRING HARBOR LAB.

PA (REGC) UNIV CALIFORNIA.

XX Bell SP, Foss M, Herskowitz I, Kobayashi R, Laurensen P;

PI Li JJ, McNally FJ, Rine J, Stillman BW;

XX WPI; 1997-052354/05.

DR N-FSDB; T62360.

XX Nucleic acid encoding origin of replication complex (ORC) protein -

PT useful to screen for lead pharmaceuticals capable of disrupting ORC

PT protein function, and inhibiting cell growth

XX Disclosure; Page 30-34; 57pp; English.

XX Origin of replication (ORC) proteins (W14136-41) are respectively

CC encoded by cDNA clones (T62358-63) from Kluyveromyces fragilis,

CC Schizosaccharomyces pombe, human (ORC1), Arabidopsis thaliana,

CC Caenorhabditis elegans and human (ORC2). The ORC polypeptides

CC can be produced in transformed host cells, and in transgenic

CC animals for functional studies (e.g. the efficacy of candidate

CC drugs for diseases associated with expression of ORC). The

CC recombinant ORC proteins can be used in a novel method of screening

CC for lead pharmaceuticals, esp. cpds. capable of disrupting ORC

CC function and inhibiting cell growth, useful in the treatment of

CC neoplasia, inflammation, hypersensitivity, etc.

XX SQ Sequence 861 AA;

Query Match 4.8%; Score 100; DB 18; Length 861;

Best Local Similarity 21.3%; Pred. No. 0.69; Mismatches 139; Indels 100; Gaps 15;

Matches 79; Conservative 53;

Qy 55 TEPLTME-----KMSNV-----VKYWTTCPS--NTVKTENATGPEELGLPLQRSYSEHLGY 103

Db 226 thltparkrlelgnlgnpqmsqqtscasldspgrikrvafseitspskrsqdklqt 285

Qy 104 FPTDLFACSELRNGLELNASLSEFEKKNKISLLH-----SSKEKLRERIKY 153

Db 286 lspalkapektretg-----lsyteddkkaspehrilrtriaaasktidiree---- 333

Qy 154 CCEQLRTLPLPVKGRKNDAAVLEATVDYVKYIREKISPAVMAQITEALQSNMRECKKQ 213

Db 334 -----rtltplsgggr-----ssvpsvilkpenikkrdaekaq-neatstphri----- 379

Qy 214 TPIELSLPGTVMAQRENSVMSTYSPERGLOFTNTCWNGCSTPDAAESSLDEAVRPSSSA 273

Db 380 -----rrkssvltmnrirqlrflgns-----ksdqeekelipaeeisdss 421

Qy 274 SENAGDPYKTHISSAALSIN-----SLHTVRYKVTSPSYDATATVNQINISHLSP-S 325

Db 422 deeeastpplpraprtvsnrlsrslkslhtl-----tkvpkkslkprtprc 469

Qy 326 AMPVVSFSFLG-----TALLGWARRALHIPTVCNS-----FGRIKSTCLKFTLST 370

Db 470 aapqirsrlaapeasvleearlhsvsavgpeslpcrqeqfdqlynfveskllhtgtgc 529

Qy 371 TYWAQFDNLGK 381

Db 530 myisgvpgtgk 540

RESULT 15

W09381

ID W09381 standard; protein; 728 AA.

XX W09381;

AC W09381;

XX 29-MAY-1997 (first entry)

XX Chicken heat shock protein-90.

DE Chicken; heat shock protein-90; HSP-90; neural spine elongation factor;

Db 1588 assvpkeeklpeadvavfrlfrqgasilsqyprllpqgaanqpldspchqasllsrr 1647

QY 293 LNSLHTVRYYSKYTPSYDANAVTNQISHLPSAMPVVSSES-----LGRA----- 338

Db 1648 whlightlrlnk-----prtmknqgssslavsspsptavafstnggraavgtangtvy1 1702

QY 339 --LLGWARRALHPIPTVNSFGRIKSTCLKFTLTSTTYWAQFDNL 379

Db 1703 ldirtw-qeeksvsvsgcdgi-----sacflsddtlflitaftdgl 1740

RESULT 12

R21261

ID R21261 standard; Protein; 241 AA.

AC R21261;

XX

DT 21-MAY-1992 (first entry)

DE VHD1.3-HUCH1 region of Fab D1.3 in pUC19.

XX

KW Fd; bacteriophage; gene III; filamentous; phagemid; capsid; coat;

KW pilus; g3p; binding; adsorption; gene VIII; diverse repertoire;

KW specific binding pairs; replicable genetic display package; Ig;

KW immunoglobulin; antibody; heavy chain; light chain; variable.

XX

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Peptide 1..22

FT /label= PelB_leader

FT Region 23..138

FT /label= VHD1.3

FT Peptide 139..241

FT /label= HUCH1

XX

PN WO9201047-A.

XX

PD 23-JAN-1992.

XX

PF 10-JUL-1991; 91WO-GB01134.

XX

PR 15-MAY-1991; 91GB-0010549.

PR 10-JUL-1990; 90GB-0015198.

PR 19-OCT-1990; 90GB-0022845.

PR 12-NOV-1990; 90GB-0024503.

PR 06-MAR-1991; 91GB-0004744.

XX

PA (CAMP-) CAMBRIDGE ANTIBODY.

PA (MEDI-) MED RES COUNCIL.

XX

PI McCafferty J, Pope AR, Johnson KS, Hoogenboom HRJ, Griffiths AD;

PI Jackson RH, Holliger KP, Marks JD, Clackson TP, Chiswell DJ;

PI Winter GP, Bonnett TP;

XX

WPI: 1992-056862/07.

N-PSDB; Q21097.

XX

PT Producing members of specific binding pairs - by expression in

PT recombinant host cells with a secreting replicable genetic

PT display package.

XX

PS Example 7; Fig 10; 209pp; English.

XX

CC The plasmid was the starting point for the prepn. of a pAb that

CC expresses a Fab-like fragment on its surface. (The antibody D1.3

CC is directed against hen egg lysozyme). The coding sequence

CC are preceded by the pelB signal peptide sequence (Ward ES et al,

CC 1989., Nature 341 544-546 (1989). The sequence encoding the VH

CC and CH1 regions shown here was amplified and ligated into f0CAR2

CC (see Q23463) which contains the gene III sequence to construct

CC f0CAR2VHCH1 D1.3. The heavy chain was deleted from Fab D1.3 in

CC pUC19 and the 2.7 kb fragment contg. the light chain was purified

CC and self-ligated to prepare LCD1.3 DHC. Supernatant contg. phage

CC particles from cells transformed with fd CAT2VHCH1 D1.3 were

CC cultured with LCD1.3 DHC cells and phage from the resulting colonies

CC were assayed for ability to bind to lysosyme. Results showed that

CC a functional Fab fragment was produced by an association of the free

CC light chain (see R22565) with the VHCH1 fragment (shown here) fused

CC to gene III and expressed on the surface of the pAb.

CC See also R21260-307, 309-311; R22450, 565-581.

XX

SQ Sequence 241 AA;

Query Match 4.8%; Score 100.5; DB 13; Length 241;

Best Local Similarity 25.7%; Pred. No. 0.069;

Matches 55; Conservative 19; Mismatches 73; Indels 67; Gaps 10;

QY 213 QTFIELSLPCTVMAQRENSVMSTYS-----PERGLQFLTWTNCGSTPD 257

Db 23 qvqlgespglvaqpsqlsictvsgfsltygvnvrppgkgle-----wlgmiwgd 76

QY 258 AESSLDEAVRPSSASANAIGDPYKTHISSAALSLSLH---TVRYYSKVTPSY----- 309

Db 77 gntdymsalkrslskdms-----ksqvlkmslhtddtaryycarerdyldyw 128

QY 310 -DATAVTNQNISIHLPASMP--PVS-SFSLGTALLG-----WARRAL----- 347

Db 129 gggtvtvssastkqpsvflapsskstggttaalgclykdyfpeptvswnsgaltsgv 188

QY 348 -HIPTVCNSEGRKSTCLKFTLTSTTYWAQFDNLG 380

Db 189 htfpavlgsg-----lyslssvvtvpsslg 215

RESULT 13

W22232

ID W22232 standard; Protein; 861 AA.

XX

AC W22232;

XX

DT 12-SEP-1997 (first entry)

DE Human origin of replication complex protein 1.

XX

KW Origin of replication complex; ORC; yeast; human; chromatography;

KW peptide sequencing; primer; amplification; PCR; genome; cell growth;

KW polymerase chain reaction; open reading frame; cancer; infection;

KW inflammation; hypersensitivity.

XX

OS Homo sapiens.

XX

PN US5614618-A.

XX

PD 25-MAR-1997.

XX

PF 16-DEC-1993; 93US-0168479.

XX

PR 07-JUN-1995; 95US-0484106.

PR 16-DEC-1993; 93US-0168479.

XX

XX (COLD-) COLD SPRING HARBOR LAB.

PA (REGC) UNIV CALIFORNIA.

XX

PI Bell SP, Foss M, Gavin K, Herskowitz I, Hidaka M;

PI Kobayashi R, Laurensen P, Li J, McNally FJ, Rine J;

PI Stillman BW;

XX

WPI: 1997-201534/18.

N-PSDB; T73287.

XX

PT Nucleic acids encoding origin of replication complex proteins - used

PT for screening for lead cpds. for therapy or diagnosis of disease

PT associated with undesirable cell growth

[illegible]

mSin nucleic acids encoding recombinant polypeptide(s) that associate with Mad polypeptide - are possible homologues of *S. cerevisiae* general repressor protein

Example 2; Fig 2D; 111pp; English.

AC	Y07085;
XX	
DT	02-JUL-1999 (first entry)
XX	
DE	Renal cancer associated antigen precursor sequence.
XX	
KW	Cancer associated antigen; diagnosis; research; treatment; human;
KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW	prostate cancer.
XX	
OS	Homo sapiens.
XX	
PN	WO904265-A2..
XX	
PD	28-JAN-1999.
XX	
PF	15-JUL-1998; 98WO-USI4679.
XX	
PR	22-JUN-1998; 98US-0102322.
PR	17-JUL-1997; 97US-0896164.
PR	10-OCT-1997; 97US-0061599.
PR	10-OCT-1997; 97US-0061765.
PR	10-OCT-1997; 97US-0948705.
PR	11-OCT-1997; 97GB-0021697.
XX	
PA	(LUDW-) LUDWIG INST CANCER RES.
XX	
PI	Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI	pfeundschuh M, Sahin U, Scanlan MJ, Stockert B;
PI	Tureci O;
DR	WPI; 1999-132448/11.
XX	
PT	New isolated cancer associated nucleic acids and polypeptides -
PT	isolated using sera from cancer patients, used to develop products
PT	for the diagnosis, monitoring or treatment of cancers
XX	
PS	Disclosure; Page 497-499; 787pp; English.
XX	
CC	The invention relates to a method for diagnosing a disorder characterised
CC	by expression of a human cancer associated antigen precursor coded for by
CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC	biological sample isolated from a subject with an agent that specifically
CC	binds to the NAM, an expression product or a fragment of an expression
CC	product complexed with an HLA molecule; and (b) determining the
CC	interaction between the agent and the NAM or the expression product as a
CC	determination of the disorder. The products and methods can be used in
CC	the diagnosis, monitoring, research, or treatment of conditions
CC	characterised by the expression of various cancer associated antigens.
CC	The invention provides nucleic acid sequences and encoded polypeptides
CC	which are cancer associated antigen precursors expressed in human breast
CC	cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC	lung cancer.
XX	
SQ	Sequence 732 AA;
	Query Match 5.4%; Score 113.5; DB 20; Length 732;
	Best Local Similarity 21.8%; Pred. NO. 0.023;
	Matches 81; Conservative 58; Mismatches 141; Indels 91; Gaps 18;
QY	6 VPSSLAELEAIAKILIRGKK---KNTHSLFVFI--PENFKGISGHGMIDIAL---TEP 57 ! : : ! : : : : ! : : : : : ! : : : : : ! : : : : :
Dd	394 lplnlsremldqgskilvirknlvkckcllftelaedkenykkfyeqfskniklgheds 453 ! : : : : : ! : : : : : ! : : : : : ! : : : : : ! : : : : :
QY	58 LTMEKMSNVVKYWTCPNTVKTENATPEELGL----PLORSYSEHLGYFFPTDL----- 108 ! : : : : : ! : : : : : ! : : : : : ! : : : : : ! : : : : :
Dd	454 qnrkksellryyt-----sasgemsvslkdycrtnkengkhyyigetkdqva 503 ! : : : : : ! : : : : : ! : : : : : ! : : : : : ! : : : : :
QY	109 -FACESLRNNGNLUEL-----NASUSEFEFNKKIS-----LLHSSKEKLRRRE 149 ! : : : : : ! : : : : : ! : : : : : ! : : : : : ! : : : : :
Dd	504 nsatverlrk-haleviymiepidvcvgqlkefegktlvsvtklelpedeekkkde 562 ! : : : : : ! : : : : : ! : : : : : ! : : : : : ! : : : : :

QY	150	RIRYCCBQLTLLPYVKGRKNDASVLEATVDYVKVIREKI-SPAYMAOITFALQSNMRF	209
	: :	: : :	:
Db	563	ektkkfenlckim-----kdilekkvkvsvsnrlvtspccivltstgygtanmer	612
QY	209	CKKOQTPIELSLPGTVMAORENSVMSTSPERGLQFLTNTCWNGCSTPDAESSLDEAVRV	268
			:
Db	613	lnkaq-----alrdnstmymaakkhlel-	668
QY	269	PS-SSASENAIGD----PYKTHISSAALSIN--SLHTVRYYS--KVTPSYDATAVTNQNI	319
	: :	: :	:
Db	649	kacadmksvkdvlllyetalallsgsfslsdpqthanrlyrmiklgldiddptadtt	708
QY	320	SIHLPSAMPPV 330	
	: :		
Db	709	saavteemppl 719	
RESULT	4		
R62656	ID	RG2656 standard; Protein; 383 AA.	
XX	AC	R62656;	
XX	DT	08-JUN-1995 (first entry)	
XX	DE	Petunia Ph6 gene product.	
XX	KW	Vacuolar pH; pH gene; Petunia.	
XX	OS	Petunia hybrida strain V26.	
XX	FH	Key Location/Qualifiers	
FT	Region	192..198	
FT	/label=	Helix-loop-helix structural motif	
FT	/note=	"see also 239..245"	
XX	WO9423561-A.		
XX	27-OCT-1994.		
XX	15-APR-1994;	94WO-US04173.	
XX	16-APR-1993;	93US-0049282.	
XX	(DNAP) DNA PLANT TECHNOLOGY CORP.		
XX	Chuck GS, Courtney-guterson N, Dooner HK, Keller J;		
PI	Nijjar CS, Ralston EJ;		
XX	WPI; 1994-341349/42.		
DR	N-PSDB; Q73000.		
XX	Petunia Ph gene and constructs containing it - for alteration of vacuolar pH used in the formation of blue flowers		
XX	Disclosure; Page 42-44; 62pp; English.		
XX	The V26 strain of petunia was used. Poly A RNA was isolated from total RNA from flower buds and used to generate a cDNA library in the vector lambda ZapII (Stratagene). The SstI to BamHI fragment at the left hand side of Ac was used to isolate ppet14-1 (contg. pH6 cDNA), the sequence for which is claimed (see Q73000 Ff). The AA sequence comprises a helix-loop-helix structural motif starting with the sequence NHVLAEK (starting at residue 192) and extending to the sequence KKYQDLK (ending at residue 245). Proteins including this motif include the myc family of oncogenes, regulators of neuron and muscle development, and regulators of segmentation and organ patterning in prosophal. In plants, the motif is found in R(S), a protein involved with regulating anthocyanin synthesis in Maize. Comparison of the pH6 and R(S) sequences over the 54 AA helix-loop-helix region detected 57% AA identity. Outside this region only 19% identity was found.		
XX			

PA (HUBE/) HUBERT R S.
 PA (RAIT/) RAITANO A B.
 XX Afar DE, Hubert RS, Raitano AB;
 XX WPI; 2000-237872/20.
 DR N-PSDB; 294275.
 XX
 PT Testis specific Helix Loop Helix proteins expressed in cancers and
 PT useful for the prevention, diagnosis and treatment of prostate, bladder
 PT and ovarian tumors -
 PS
 PS Claim 1; Fig 2A-D; 62pp; English.
 XX
 XX This sequence is that of human PHELIX, a novel basic Helix Loop
 CC Helix protein thought to act as a transcription factor. PHELIX
 CC normally exhibits a testis-specific expression pattern but is
 CC up-regulated in prostate and other types of cancer. The invention
 CC provides diagnostic and therapeutic methods useful in the
 CC management of various cancers which express PHELIX, including
 CC prostate cancer, bladder cancer, ovarian cancer and testicular
 CC cancer, including therapies aimed at inhibition the transcription,
 CC translation, processing or function of PHELIX. The expression
 CC pattern of PHELIX suggests that is an ideal target for a cancer
 CC vaccine approach to prostate cancer. PHELIX protein can also be
 CC used to screen for agonists and antagonists of therapeutic value
 CC and to raise antibodies.
 XX
 XX Sequence 405 AA;
 SQ
 Query Match 100.0%; Score 2095; DB 21; Length 405;
 Best Local Similarity 100.0%; Pred. No. 6 4e-203;
 Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MVLKVPSSLSAELEAKLIRFGKKNTSLFVFIIPENFKGICSGHGMIALTEPLTM 60
 DB 1 MVLKVPSSLSAELEAKLIRFGKKNTSLFVFIIPENFKGICSGHGMIALTEPLTM 60
 QY 61 EKMSNVVYKWTCTSNVKTENATGPEELGLPLQRSYSEHLGYPTDLFACSESIRNGNG 120
 DB 61 EKMSNVVYKWTCTSNVKTENATGPEELGLPLQRSYSEHLGYPTDLFACSESIRNGNG 120
 QY 121 LELNASLSEFEKKNKISLLSHSSEKLRERIKYCEQLRTLPLPVYKGNDAASVLEATV 180
 DB 121 LELNASLSEFEKKNKISLLSHSSEKLRERIKYCEQLRTLPLPVYKGNDAASVLEATV 180
 QY 181 DYVKYIREKISPAVMAOITEALQSNMRFCKKQOQTPIELSLPGTYMAQRENSVMSTYSPE 240
 DB 181 DYVKYIREKISPAVMAOITEALQSNMRFCKKQOQTPIELSLPGTYMAQRENSVMSTYSPE 240
 QY 241 GLOFLTNTCKNGCSTPDAESLDEAVRPSSASENAGDPYKTHISSAALSLSLHTVR 300
 DB 241 GLOFLTNTCKNGCSTPDAESLDEAVRPSSASENAGDPYKTHISSAALSLSLHTVR 300
 QY 301 YYSKVTSPYDATAYTNONISIHLPSPMPYPVSSFSGLGTALLGWARRALHIPTVCNSFGRIK 360
 DB 301 YYSKVTSPYDATAYTNONISIHLPSPMPYPVSSFSGLGTALLGWARRALHIPTVCNSFGRIK 360
 QY 361 STCKFLTSTTYWAQFNLGKVEORMILKAPPKDLISKELAWFGF 405
 DB 361 STCKFLTSTTYWAQFNLGKVEORMILKAPPKDLISKELAWFGF 405
 RESULT 2
 Y67501
 ID Y67501 standard; Protein; 158 AA.
 XX Y67501;
 XX
 XX 19-MAY-2000 (first entry)
 XX PBR-associated protein (PAP)8.
 DE

XX
 KW Peripheral-type benzodiazepine receptor; PBR; PBR- associated protein;
 KW PAP; cell proliferation; cancer; cell death; cytostatic; neuroprotective;
 KW immunomodulator; antinfertility; cerebroprotective; atherosclerosis;
 KW Niemann-Pick C; tumour; Alzheimer's disease; developmental disorder;
 KW cholesterol; multiple sclerosis; stress; neurodegenerative disorder;
 KW immune disorder; stroke; PAP8.
 XX
 OS Mus sp.
 XX
 XX WO200009549-A2.
 XX
 XX 24-FEB-2000.
 XX
 XX 11-AUG-1999; 99WO-US18507.
 XX
 XX 11-AUG-1998; 98US-0096048.
 XX
 XX (GEOU) UNIV GEORGETOWN MEDICAL CENT.
 XX
 XX Papadopoulos V, Li H;
 XX WPI; 2000-224278/19.
 DR N-PSDB; 257039.
 DR
 XX Novel peripheral-type benzodiazepine receptor associated proteins used
 PT for the regulation of the peripheral-type benzodiazepine receptor -
 XX
 XX Claim 20; Page 70; 71pp; English.
 PS
 XX The invention provides isolated peripheral-type benzodiazepine receptor
 CC (PBR) associated proteins (PAPs) and nucleic acids encoding the PAP
 CC proteins. The PAP polynucleotides are a source of primers and probes for
 CC detection, isolation and amplification. PAP ligands or substrates or
 CC antibodies can be labeled and used to detect PAPs, in the diagnosis and
 CC prognosis of disease associated with increased cell proliferation, such
 CC as cancer, or reduced cell death. The diagnostic methods of the
 CC invention can be predictive of diseases involving PBR including
 CC gallstones, atherosclerosis, Niemann-Pick C, Sitosterolemia, Dystrophy,
 CC tumor proliferation, Schnyder's corneal crystalline dystrophy, brain
 CC disorders including Alzheimer's disease, cholesterol metabolism,
 CC Tellurium toxicity, Smith-Lemli-Opitz syndrome, myelinization,
 CC developmental abnormalities, demyelination, Charcot-Marie tooth
 CC disease, Pelizaeus-Merzbacher disease, Multiple sclerosis, and SLA. The
 CC methods may also be useful in prophylactic treatments, or in screening
 CC for compounds effective in prophylactic treatment. The PAPs may be
 CC used to identify inhibitors or activators which allows the identification
 CC of drugs or agents which modulate PBR activity. Inhibitors of PAP may be
 CC used in the treatment or amelioration of conditions such as stress and
 CC stroke, cancer, neurodegenerative disorders, developmental disorders,
 CC infertility and immune disorders. The present sequence represents
 CC a PAP8 polypeptide.
 XX
 XX Sequence 158 AA;
 SQ
 Query Match 5.5%; Score 115; DB 21; Length 158;
 Best Local Similarity 27.7%; Pred. No. 0.0011;
 Matches 28; Conservative 24; Mismatches 41; Indels 8; Gaps 3;
 QY 140 HSSKEKLRRERIKYCCQQLTLLPVYKGNDAASVLEATVYVKYIREKISPAVMAOIT 199
 DB 140 HSSKEKLRRERIKYCCQQLTLLPVYKGNDAASVLEATVYVKYIREKISPAVMAOIT 199
 QY 200 EALQSNMRFCKKQOQTPIELSLP-GTYMAQRENSVMSTYSPE 239
 DB 200 EALQSNMRFCKKQOQTPIELSLP-GTYMAQRENSVMSTYSPE 239
 QY 122 sv-----fcgktrrlkltrpslvtcpgagsgspame 156
 DB 122 sv-----fcgktrrlkltrpslvtcpgagsgspame 156
 RESULT 3
 Y07085
 ID Y07085 standard; Protein; 732 AA.
 XX

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 19, 2001, 01:23:10 ; Search time 69.34 Seconds
(without alignments)
199.719 Million cell updates/sec

Title: US-09-389-000-2
Perfect score: 2095
Sequence: 1 MVLKVPSSLSAELEAIKL.....MILKAPPKDLISKELAWGFG 405

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_35:*

1: /SID56/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SID56/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SID56/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SID56/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SID56/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SID56/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SID56/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SID56/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SID56/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SID56/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SID56/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SID56/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SID56/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SID56/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SID56/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SID56/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SID56/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SID56/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SID56/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SID56/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SID56/gcgdata/geneseq/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2095	100.0	405	21	Human testis-speci
2	115	5.5	158	21	PBR-associated pro
3	113.5	5.4	732	20	Renal cancer assoc
4	108.5	5.2	383	15	Petunia Ph6 gene p
5	108	5.2	151	18	Max 14. Homo sapi
6	108	5.2	160	18	Max 11. Homo sapi
7	108	5.2	1187	16	AF-4 protein (enco
8	108	5.2	1210	16	AF-4 protein (enco
9	105	5.0	96	20	Human lung tumour
10	103	4.9	160	14	Max protein. R33
11	103	4.9	2627	19	Human telomerase R
12	100.5	4.8	241	13	VHDL3-HUCHI regio

13	100	4.8	861	18	W22232	Human origin of re
14	100	4.8	861	18	W14138	Human origin of re
15	99.5	4.7	728	17	W09381	Chicken heat shock
16	97.5	4.7	534	20	W82799	Human IRS-p53h pro
17	97.5	4.7	1289	20	Y27163	Peptide Seq ID No:
18	97.5	4.7	1289	21	Y56781	Human Trad protein
19	97.5	4.7	1704	15	R49657	Sequence Of Heliot
20	96	4.6	412	18	W08140	Human cytokine res
21	96	4.6	436	19	W69509	Mouse Pax6 protein
22	96	4.6	436	19	W69513	Mouse Pax6 protein
23	96	4.6	436	20	Y32831	Pax6 protein sequ
24	96	4.6	945	20	Y35612	C. pneumoniae prot
25	95	4.5	363	21	W90763	A. aeolicus AAEOL3
26	95	4.5	363	21	Y52035	A. aeolicus AAEOL3
27	95	4.5	363	21	Y51664	A. aeolicus AAEOL3
28	94	4.5	323	20	Y06112	Nelson Bay virus s
29	94	4.5	412	20	W99060	Human basic helix-
30	92.5	4.4	1254	11	R07503	Meroprote apical-e
31	92.5	4.4	1254	18	W24375	Meroprote apical-e
32	92	4.4	734	13	R20210	Diacylglycerol k1
33	92	4.4	992	20	W96265	Murine axin. Mus
34	92	4.4	2625	19	W55887	Human telomerase.
35	91	4.3	221	14	R33387	Max protein. R33
36	91	4.3	221	18	W10043	Human Mad-1. Homo
37	91	4.3	768	17	W00811	Ras p21 interactin
38	91	4.3	768	20	Y31244	Human RGL protein.
39	91	4.3	2308	15	R52580	RPTP-beta amino ac
40	91	4.3	2308	15	R57902	Human RPTP-beta.
41	90.5	4.3	551	20	Y73938	Human prostate tum
42	90.5	4.3	1041	20	W30613	Human CDC28-#3 RNA
43	90.5	4.3	1142	21	Y43876	Amino acid sequenc
44	89.5	4.3	425	20	Y24069	Salmonella typhimu
45	89.5	4.3	928	11	R06289	Predicted retinobl

ALIGNMENTS

RESULT 1

Y79269
ID Y79269 standard; Protein; 405 AA.

XX

AC Y79269;

XX

DT 03-JUL-2000 (first entry)

XX

DE Human testis-specific transcription factor PHELIX.

XX

KW PHELIX: human; testis-specific; transcription factor;

KW prostate cancer; bladder cancer; ovary cancer; testicular cancer;

KW therapy; diagnosis; vaccine.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Peptide 134..150

FT Peptide /note= "nuclear localization signal"

FT Peptide 163..169

FT Peptide /note= "nuclear localization signal"

FT Domain 140..189

FT /note= "basic Helix-Loop-Helix domain"

XX

XX WO200012709-A2.

XX

XX 09-MAR-2000.

XX

XX 31-AUG-1999; 99WO-US20137.

XX

XX 31-AUG-1998; 98US-0098610.

XX

XX 31-OCT-1998; 98US-0106524.

XX

XX (UROC*) UROGENESYS INC.

XX (AFAR*) AFAR D E.

QY 58 LTMKMSNVK YWITCPSNT

Db 454 ONRKLSELRLRYT-----SASGDEMYSLKDYCTRMKNCKHIYITGETKDYA 503
Qy 109 -FACESLJNGLNEL-----NASLSEFEKNKKIS-----LLHSSKEKLRE 149
Db 504 NSAFVERLKR-HGLEIYMIIEPIDEXCYOOLKEFEKGTIVSYTKESGLELPEDEEEKKOE 562
Qy 150 RIKYCEOLRTLLPYKGRKNDAAVLEATVDYVYKIRREKI-SPNAMAITEALGNSMNF 208
Db 563 EKKTYFENLCKTM-----KDLLEKKVEKVVVSNRLVSPCLVSTYGTWITNMR 612
Qy 209 CKKOOTPIELSLPGYVMAORENSVMSYSPERGLOFLTNTCWNGCSTPDASSLDEAVY 268
Db 613 IKKAQ-----ARONSTMGYMAKKHLEI-----NPD--HSIIETLRQ 648
Qy 269 PS-SASENAIGD---PKTHISSALSLN--SLHTVYYS--KTPSYDATATYTNONI 319
Db 649 KAEAKNKNSVDLVLLYETALLSGFSLDDPQHNRIRYRKIGLIGIDDDPDADYT 708
Qy 320 SIHLPSAMPV 330
Db 709 SAAVTEEMPL 719

RESULT 2

US-08-049-282B-2
Sequence 2, Application US/08049282B

Patent No. 5534660
GENERAL INFORMATION:
APPLICANT: CHUCK, George S.
APPLICANT: DOONER, Hugo K.
APPLICANT: COURTNEY-GUTTERSON, Neal
APPLICANT: KELLER, Janis
APPLICANT: NIJJAR, Charanjit S.
APPLICANT: RALSTON, Edward J.
TITLE OF INVENTION: PH GENES AND THEIR USES
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Hourie and Crew
STREET: Stewart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/049,282B
FILING DATE: 16-APR-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 12176-34
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-049-282B-2

Query Match 5.2%; Score 108.5; DB 1; Length 382;
Best Local Similarity 23.5%; Pred. No. 0.01; 113; Indels 75; Gaps 15;
Matches 72; Conservative 46; Mismatches 113; Indels 75; Gaps 15;
Qy 6 VPSISASELEIAIKIRGKKRKNH-SLFLVILIPENFK-----GCISGHGMIDA 53

Db 63 LPGPSYDEL-----SQEDTHYSQTVSTILEHLNOSKBSSTIMGCS----- 106
Qy 54 LPEPLTMEKMSNVYKWTCPSNVTKENATGPEELGLPLOSYSHEHLYGPTDLFACSE 113
Db 107 -----QTQSAFTRWPS-PSTVSSPFDGAGNSGOMLKSILFVSPFLHTKYQTAAE 158
Qy 114 -SLRNGNLELNAS-LSEFEKNKI-----SLHSSKEKLREKIRKICYCEOLRTLLPYK 166
Db 159 VSPKSRDATYVDSSTASFRKGCSTIOEPPSGNHYLAERRRREKINERFIILRSLVPEV- 217
Qy 167 GRKNDAAVLEATVDYVYKIRKISP-AVMAOTPEALGNSMRFCKKOOTPIELSLPGYV 225
Db 218 -TKMDKASILDGTITFYVQOLKRYQDLERANQTA-----TLQTKDTGTVK 263
Qy 226 AORENSVMSYSPERGLOFLTNTCWNG-----CSTPDASSLDEAVRPSSASENAIGD 280
Db 264 VLQGRG-----KRRMKVIGSVGGQAKITASSPTTHE-ELIVQVEVSIIESDALVE 315
Qy 281 ---PYK 283
Db 316 LRCPYK 321

RESULT 3

US-08-537-715-2
Sequence 2, Application US/08537715

Patent No. 5810627
GENERAL INFORMATION:
APPLICANT: Chuck, George S.
APPLICANT: Dooner, Hugo K.
APPLICANT: Courtney-Guterson, Neal
APPLICANT: Keller, Janis
APPLICANT: Nijjar, Charanjit S.
APPLICANT: Ralston, Edward J.
TITLE OF INVENTION: PH Genes and their uses
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,715
FILING DATE: 16-FEB-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/049,282
FILING DATE: 16-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US94/04173
FILING DATE: 15-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 012176-003410US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-537-715-2

Query Match 5.2%; Score 108.5; DB 2; Length 382;
Best Local Similarity 23.5%; Pred. No. 0.01;
Matches 72; Conservative 46; Mismatches 113; Indels 75; Gaps 15;

QY 6 VPSSLSAELEIAKILIRGKKKNT-H-SLFVFIIPENFK-----GCISGHGMDIA 53
DB LPGGSPYDEL-----SQEDTHYSQTVSTILLEHLSNOSKFSSTIMGICIS-----106

QY 54 LTFPLTKMSNVVYKWTCPSTNTVKTENATGPBELLGLPLOSSEHLGIFPTDLFACSE 113
DB 159 VSPKSRDATYDSSSTASFRKGCSTIOEPPSGNHVLAERRRERKLNERTILRLSLVPEV- 217

QY 167 GRKNDASVLEATVDYVYIREKISP-AVMAQITEALOSNMRFCKKQOTPIELSLPTVM 225
DB 218 -TKMDKASILGDTIEYVQLRKKVODLEARANOEA-----TLQTKDTGVK 263

QY 226 AORENSVSTSPERGLOFLTNCNG-----CSTPDAESSIDEAVRPSSASENAIGD 280
DB 264 VLOGRG-----KRRMKIVEGSGVGGQAKITASSPTTHE-ELIYQVEVSIIESDALVE 315

QY 281 --PYK 283
DB 316 LRCPYK 321

RESULT 4
PCT-US94-04173-2
Sequence 2, Application PC/TUS9404173
GENERAL INFORMATION:
APPLICANT: PH GENES AND THEIR USES
TITLE OF INVENTION: 4
NUMBER OF SEQUENCES: 4
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04173
FILING DATE: 15-APR-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/049,282
FILING DATE: 16-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 12176-34-1PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-04173-2

Query Match 5.2%; Score 108.5; DB 4; Length 382;
Best Local Similarity 23.5%; Pred. No. 0.01;
Matches 72; Conservative 46; Mismatches 113; Indels 75; Gaps 15;

QY 6 VPSSLSAELEIAKILIRGKKKNT-H-SLFVFIIPENFK-----GCISGHGMDIA 53
DB LPGGSPYDEL-----SQEDTHYSQTVSTILLEHLSNOSKFSSTIMGICIS-----106

QY 54 LTFPLTKMSNVVYKWTCPSTNTVKTENATGPBELLGLPLOSSEHLGIFPTDLFACSE 113
DB 107 -----QTOSAFTRMPS-PSTTVSSPFLDGATSGQMLKSLFVSFPLHTKYQTAAE 158

QY 114 -SLRNGNLELNAS-LSFEENKKI-----SLHSSKRLRERIKYCCEDLRLTLPLPYK 166
DB 159 VSPKSRDATYDSSSTASFRKGCSTIOEPPSGNHVLAERRRERKLNERTILRLSLVPEV- 217

QY 167 GRKNDASVLEATVDYVYIREKISP-AVMAQITEALOSNMRFCKKQOTPIELSLPTVM 225
DB 218 -TKMDKASILGDTIEYVQLRKKVODLEARANOEA-----TLQTKDTGVK 263

QY 226 AORENSVSTSPERGLOFLTNCNG-----CSTPDAESSIDEAVRPSSASENAIGD 280
DB 264 VLOGRG-----KRRMKIVEGSGVGGQAKITASSPTTHE-ELIYQVEVSIIESDALVE 315

QY 281 --PYK 283
DB 316 LRCPYK 321

RESULT 5
US-08-537-715-4
Sequence 4, Application US/08537715
Patent No. 5910627
GENERAL INFORMATION:
APPLICANT: Chuck, George S.
APPLICANT: Dooner, Hugo K.
APPLICANT: Courtney-Gutteson, Neal
APPLICANT: Keller, Janis
APPLICANT: Nijjar, Charanjit S.
APPLICANT: Ralston, Edward J.
TITLE OF INVENTION: PH Genes and Their Uses
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,715
FILING DATE: 16-FEB-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/049,282
FILING DATE: 16-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US94/04173
FILING DATE: 15-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 012176-003410US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 383 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-537-715-4

```

Query Match          5.2%; Score 108.5; DB 2; Length 383;
Best Local Similarity 23.5%; Pred. No. 0.01;
Matches 72; Conservative 46; Mismatches 113; Indels 75; Gaps 15;

QY 6 VPSSLSAELEAIKILRFCKKNTK-SLFVFIIPENK-----GCISGHGMDIA 53
DB 64 LPGSPSYDEL-----SQEDTHYSQVSTLLEHLNSQSKFSSITMGCIS----- 107
QY 54 LTPPLMEKSNVVKYWTCPSTVTKTENATGPELGPQRYSSEHLGFPFDLFCASE 113
DB 108 -----QTQSAFTWPS-PSTTVSSPFLDGGATSGQWLKLSLFSVPLHARKYQTAAE 199
QY 114 -SLRNGNGLELNAS-ISEEKNKKI-----SLHSKSKELRERIKYCCQQLRTLLPYVK 166
DB 160 VSPKSDATVDSSTASRFKSGSITQEEPSGNHVAERRRERKNERFIILSLVPEV- 218
QY 167 GRKNDASVLEATVDYVKYIREKISP-AVMAQITTEALQSNMFCCKQCPPIELSLPGTVM 225
DB 219 -TKMDKASITLGDITIEYVKQLRRKVDLEARNQTEA-----TLQTKDTGTVK 264
QY 226 AORENSVMSTYSPERGLQFLTNTCWNG-----CSTPAESSLDEAVRPSSASANAIGD 280
DB 265 VLAGRG-----KRRMKIVESVGGQAKITASSPSTTHE-EEIVQEVSIIESDALVE 316
QY 281 ---PYK 283
DB 317 LRCPYK 322

RESULT 6
PCT-US94-04173-4
; Sequence 4, Application PC/TUS9404173
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: PH GENES AND THEIR USES
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04173
; FILING DATE: 15-APR-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/049,282
; FILING DATE: 16-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 12176-34-1PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 383 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-04173-4

Query Match          5.2%; Score 108.5; DB 4; Length 383;
Best Local Similarity 23.5%; Pred. No. 0.01;
Matches 72; Conservative 46; Mismatches 113; Indels 75; Gaps 15;

QY 6 VPSSLSAELEAIKILRFCKKNTK-SLFVFIIPENK-----GCISGHGMDIA 53
DB 64 LPGSPSYDEL-----SQEDTHYSQVSTLLEHLNSQSKFSSITMGCIS----- 107
QY 54 LTPPLMEKSNVVKYWTCPSTVTKTENATGPELGPQRYSSEHLGFPFDLFCASE 113
DB 108 -----QTQSAFTWPS-PSTTVSSPFLDGGATSGQWLKLSLFSVPLHARKYQTAAE 199
QY 114 -SLRNGNGLELNAS-ISEEKNKKI-----SLHSKSKELRERIKYCCQQLRTLLPYVK 166
DB 160 VSPKSDATVDSSTASRFKSGSITQEEPSGNHVAERRRERKNERFIILSLVPEV- 218
QY 167 GRKNDASVLEATVDYVKYIREKISP-AVMAQITTEALQSNMFCCKQCPPIELSLPGTVM 225
DB 219 -TKMDKASITLGDITIEYVKQLRRKVDLEARNQTEA-----TLQTKDTGTVK 264
QY 226 AORENSVMSTYSPERGLQFLTNTCWNG-----CSTPAESSLDEAVRPSSASANAIGD 280
DB 265 VLAGRG-----KRRMKIVESVGGQAKITASSPSTTHE-EEIVQEVSIIESDALVE 316
QY 281 ---PYK 283
DB 317 LRCPYK 322

```

```

DB 108 -----QTQSAFTWPS-PSTTVSSPFLDGGATSGQWLKLSLFSVPLHARKYQTAAE 159
QY 114 -SLRNGNGLELNAS-ISEEKNKKI-----SLHSKSKELRERIKYCCQQLRTLLPYVK 166
DB 160 VSPKSDATVDSSTASRFKSGSITQEEPSGNHVAERRRERKNERFIILSLVPEV- 218
QY 167 GRKNDASVLEATVDYVKYIREKISP-AVMAQITTEALQSNMFCCKQCPPIELSLPGTVM 225
DB 219 -TKMDKASITLGDITIEYVKQLRRKVDLEARNQTEA-----TLQTKDTGTVK 264
QY 226 AORENSVMSTYSPERGLQFLTNTCWNG-----CSTPAESSLDEAVRPSSASANAIGD 280
DB 265 VLAGRG-----KRRMKIVESVGGQAKITASSPSTTHE-EEIVQEVSIIESDALVE 316
QY 281 ---PYK 283
DB 317 LRCPYK 322

RESULT 7
US-07-903-710-4
; Sequence 4, Application US/07903710
; Patent No. 5302519
; GENERAL INFORMATION:
; APPLICANT: Eyer, D.E.; Eisenman, R.N.; Blackwood, E.M.; Averb, D.M.
; TITLE OF INVENTION: MAX: A HELIX-LOOP-HELIX ZIPPER PROTEIN THAT FORMS A
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-5.25 inch, 1.2mb storage
; OPERATING SYSTEM: MS-DOS 4.01
; SOFTWARE: Word for Windows-1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/903,710
; FILING DATE: 19920623
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/756,195
; FILING DATE: 09/09/91
; ATTORNEY/AGENT INFORMATION:
; NAME: Sundemo, John, S.
; REGISTRATION NUMBER: 34,445
; REFERENCE/DOCKET NUMBER: FHC-1-6550
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0727 (direct)
; TELEFAX: 1-206-224-0779
; TELEX: 4938023
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
; DESCRIPTION: Human helix-loop-helix zipper protein; amino acid sequence
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens; Eukaryota; Animalia; Metazoa;
; ORGANISM: Chordata; Vertebrata; Mammalia; Theria; Eutheria;
; ORGANISM: Primates; Haplorhini; Catarrhini; Hominoidea;
; IMMEDIATE SOURCE: Human lymphoid B cell Manca cell line
; US-07-903-710-4

Query Match          5.2%; Score 108; DB 1; Length 151;
Best Local Similarity 24.4%; Pred. No. 0.0023;

```

QY 200 EALQSNMRECKKQTPIELSLF

Db 70 HTHQDIDDLKRON

GENERAL INFORMATION:
APPLICANT: Croce, Carlo
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &
ADDRESS: No. 60401401is
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545,860D
FILING DATE: 07-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04496
FILING DATE: 22-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10930
FILING DATE: 09-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/327,392
FILING DATE: 19-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/320,559
FILING DATE: 11-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,443
FILING DATE: 14-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,094
FILING DATE: 30-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/888,839
FILING DATE: 27-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/805,093
FILING DATE: 11-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Deluca Esq., Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-1262
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1187 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-545-860D-28

Query Match 5.2%; Score 108; DB 3; Length 1187;
Best Local Similarity 21.4%; Pred. No. 0.079;
Matches 64; Conservative 44; Mismatches 119; Indels 72; Gaps 12;

QY 72 TCPSNTYKTEMAT--GPEELGLPLQ-----RSTSEHIGY-----FPTDLFACSES 114
DB 868 TCGODPPKASSTKSNKDDSSIPKORVEKGSRSSSEHKGSSSDTANPFV-----PS 921
QY 115 LRNGNGLELNASLSEFEKKNKISILH-----SSKEKLRRERIKYCCEQLRTLLPYVK- 166

DB 922 LPNGNS---KPGKPYQVAFEDQADLHMRKAKKMKQKAMELTDYRGKAFKYLEAVLSTFEC 978
QY 167 -----GKKNDAASVLEATVDYVKI-----REKISPAVMAQITEALQSN 205
DB 979 GIATESSESOSKSAVSYSSEVLDIKRIMSLSKFSDATAPTOEKIFVLCMRQOSILNMA 1038
QY 206 MRFCCKQOTPIELSLPQTVMAQRENSVMSTYSPERGQFLNLCWNCGCTPDDESSIDEA 265
DB 1039 MERCKKD---IAIKYSTLKNHEBSSSKVAQAP-----SPCIASGTGTPSPISPMPSP 1087
QY 266 VRYPSSASENALGDPKYTHISSAALSINSLHT--VRYKRVTPSYD---AAVTNQN 318
DB 1088 ASSVGSQSASAGSVSSGVATISPTVITQNMTSYVITTSVILTAFLPMQAEALTRKN 1146

RESULT 15
PCT-US94-04496-28
Sequence 28, Application PC/TUS9404496
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &
ADDRESS: Noritis
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04496
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Deluca Esq., Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-1242
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1187 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-04496-28

Query Match 5.2%; Score 108; DB 4; Length 1187;
Best Local Similarity 21.4%; Pred. No. 0.079;
Matches 64; Conservative 44; Mismatches 119; Indels 72; Gaps 12;

QY 72 TCPSNTYKTEMAT--GPEELGLPLQ-----RSTSEHIGY-----FPTDLFACSES 114
DB 868 TCGODPPKASSTKSNKDDSSIPKORVEKGSRSSSEHKGSSSDTANPFV-----PS 921
QY 115 LRNGNGLELNASLSEFEKKNKISILH-----SSKEKLRRERIKYCCEQLRTLLPYVK- 166
DB 922 LPNGNS---KPGKPYQVAFEDQADLHMRKAKKMKQKAMELTDYRGKAFKYLEAVLSTFEC 978
QY 167 -----GKKNDAASVLEATVDYVKI-----REKISPAVMAQITEALQSN 205

```

Db 979 GIATESQSASVSVSEVVDLIKFIIMSLKSPSDATAPTOEKIFAVLCMRCSILNMA 1038
QY 206 MRCKKOQTPLESLPGTVMAQRENSVMSSTYSPERGLOFLTNTCWNCGSTPDAESIDEA 265
Db 1039 MFRCKD---IAIKYSRTLKHFESSKVAQAP-----SPCIASTGTPTSPPLSPMPS 1087
QY 266 VRPSSASASENADPYKTHISSAALSLSLHT--VRYKYVTPSYD---ATAVTNON 318
Db 1088 ASSVGSOSASAGSVSSGVAATISTPVTIQNMSTSYVTITSHVLTAFDLMEQAEALTRKN 1146

```

Search completed: January 19, 2001, 10:44:57
 Job time: 32491 sec

THIS PAGE BLANK (USPTO)

Align seg 1/1 to: AK000456 from: 1 to: 2147

```

1 MetValLeuLeuLysValProSerSerLeuSerAlaGluGluLeuGluAl 17
|||||
212 ATGTTCTCTTGAAGGTCCTCTCTCAAGTGCAGAGGAGCTGGAAGC 261
17 ailelyslleuileargpheglylyslysasnthrhiserleuphev 34
|||||
262 CATCAAGTTAATTAGATTGGCAAAAAGAAAATACACATTCACCTGTTG 311
34 alpheielleplogluasnphelysglycysileserghylhslglymet 50
|||||
312 TTTTATTAATCCCTGAAATTTAAAGTGTTATTCAGGGCATGAGATG 361
51 AspIleAlaLeuThrGluProLeuThrMetGluLysMetSerAsnValVa 67
|||||
362 GATATTGCTTAAGTACGACACATGACAAATGAAAAATGAGTATGCGT 411
67 LysTyrTrpThrThrCysProSerAsnThrValLysThrGluAsnAlaT 84
|||||
412 AAAATACGTGACAAACATGTCCTCAACACTGTTAAGACTGAAAAAGCAA 461
84 hrcglyProGluGluLeuGlyLeuProLeuGlnArgSerTyrSerGluHls 100
|||||
462 CTGGGCTGAGAACTTGGATTGCCCTGCAGAGGTCCTACAGCGAAAC 511
101 LeuGlyTyrPheProThrAspLeuPheAlaCysSerGluSerLeuArg 117
|||||
512 CTGGGATATTCTTCTACTGATCTATTGCTGCTGCAATCTTTAAGGAA 561
117 nglyAsnGlyLeuGluLeuAsnAlaSerLeuSerGluPheGluLysAsn 134
|||||
562 TGGCAATGGCTTGAATTAATGCTTCGTTGTCAGAGTTGCAAAAAACA 611
134 yslYslIleSerLeuLeuHlsSerSerLysGluLysLeuArgArgGluArg 150
|||||
612 AAAAGATCTCTCTTCTTCAATTCAGCAAGAAAACTAAGAGGAAAGA 661
151 IleLysTyrCysCysGluGluLeuArgThrLeuLeuProTyrValLysG 167
|||||
662 ATCAAAATATTGCTGTGAGCAGCTGCTACTCTCTCCGATGTAAAGG 711
167 YArgLysAsnAspAlaAlaSerValLeuGluAlaThrValAspTyrVal 184
|||||
712 GAGAAAGATGATGCGGCTTCACTTCTGAGGCAACAGTTGATTGCTGA 761
184 ystYrIleArgGluLysIleSerProAlaValMetAlaGlnIleThrGlu 200
|||||
762 AATATATCCGGAGAAAATCTCTCCAGCGTTATGGCCCGATTCAGAA 811
201 AlaLeuGlnSerAsnMetArgPheCysLysLysGlnGlnThrProIleG 217
|||||
812 GCACCTCAAGCAACATGAGGTTTGTAAAGAAACAACAAACCCATTTGA 861
217 uleuSerLeuProGlyThrValMetAlaGlnArgGluAsnSerValMet 234
|||||
862 GCTGTCTCTCCAGGACCTGATGSCACAGCGGAAAAACAGTGATGA 911
234 eThrTyrSerProGluArgGlyLeuGlnPheLeuThrAsnThrCysTrp 250
|||||
912 GCACCTACCCCTGAGAGAGGCTCCCAATTCCTGCTAATACGCTCTGG 961
251 AsnGlyCysSerThrProAspAlaGluSerSerLeuAspGluAlaValAr 267
|||||
962 AATGGGTGCTCCACTCTGATGAGAGACTCTCTGATGAGAACCTGTGAG 1011
267 gValProSerSerSerAlaSerGluAsnAlaIleGlyAspProTyrLysT 284
|||||
1012 AGTTCCATTAAGCTCGGCTCAGAGAAATCTATTGTTGATTCATATTA 1061
284 hrlhslIleSerSerAlaAlaLeuSerLeuAsnSerLeuHlsThrValArg 300
|||||

```

```

1062 CTCACATTTCCAGTGCAGGCGTGTCTCTGAATTCCTTCGACTACTGCAGA 1111
301 TyrTyrSerLysValThrProSerTyrAspAlaThrAlaValThrAsnG 317
|||||
1112 TATTATTCTAAAGTACACCCCTTCTCAAGTGCAGACTGCTGTAACAAATCA 1161
317 nasnlIleSerIleHlsLeuProSerAlaMetProProValSer.Serpe 333
|||||
1162 GAACATTTCAATTCAATTACTTTCAGCCATGCCCCCGGTCTGAAAGCTTC 1211
334 SerLeuGlyThrAlaLeuLeuGlyThrAlaArgArgAlaLeuHlsIlePr 350
|||||
1212 TCCCTGCGTACGCACTTCTGGGTTGGGCCAGACGTCACATACATGCC 1261
350 oThrValCysAsnSerPheGlyArgIleLysSerThrCysLeuLysPheT 367
|||||
1262 AACTGCTCTCAACAGTTTGGCGGTATTAAAAGCACATGTTTGAATTTCA 1311
367 hrlLeuSerThrThrTyrTrpAlaGlnPheAspAsnLeuGlyLysValGlu 383
|||||
1312 CACTCTCAACCACTTACTGGGCGCAGTTTGACAAATCTGAAAAAGCTGGA 1361
384 GlnArgMetIleLeuLysAlaProProLysAspLeuIleSerLysGluLe 400
|||||
1362 CAAAGATGATTTTGAAGCTCTCACCCAAAGCACTAATATCAAAAGAGTT 1411
400 uAlaTrpPheGlyPhe 405
|||||
1412 GGCATGTTGCTTC 1427

```

seq_name: gb_hlg20:AL160392

seq_documentation_block:
 LOCUS AL160392 158888 bp DNA HNG 16-SEP-2000
 DEFINITION Homo sapiens chromosome 13 clone RP11-121N13, *** SEQUENCING IN
 PROGRAMS ***, 2 unordered pieces.

ACCESSION AL160392
 VERSION AL160392.10 GI:10185503
 KEYWORDS HNG; HTGS_PHRASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 DIRECT Submission
 Submitted (15-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Sep 18, 2000 this sequence version replaced gi:10039688.

COMMENT
 ----- Genome Center

Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk
 ----- Project Information
 Center project name: BA121N13
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 156753 bases at least Q40
 Consensus quality: 156783 bases at least Q30
 Consensus quality: 156785 bases at least Q20
 Insert size: 158788; sum-of-contigs
 Insert size: 146321; 14.5% error; agarose-fp
 Quality coverage: 8.48x in Q20 bases; sum-of-contigs Quality
 coverage: 9.34x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 156788: contig of 156788 bp in length
* 156789 156888: gap of 100 bp
* 156889 158888: contig of 2000 bp in length.
Location/Qualifiers
1. 158888
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-121N13"
/clone_lib="RPC1-11.1"
1. 156788
/note="assembly_fragment:03393"

FEATURES

source

misc_feature

clone_end:SP6
vector_side:right
clone_end:T7
vector_side:left
156889. 158888
misc_feature /note="assembly_fragment:03390"
BASE COUNT 51320 a 30978 c 30194 g 46296 t 100 others
ORIGIN

alignment_scores:

Quality: 443.50 Length: 666
Ratio: 2.899 Gaps: 4
Percent Similarity: 22.973 Percent Identity: 21.321

alignment_block:

US-09-389-000-2 x AL160392/rev ..

Align seg 1/1 to reverse of: AL160392 from: 1 to: 158888

243 GlnPheLeuThrsanthrCysTrpAsnGlyCysSerThrProAspPala1 259
|||||:|||||
56866 CACTTATGTCGAAT.....AACAGTCG 56844
259 userSerLeuAspGluAa.....V 266
:|||||:|||||
56843 TTCGTAAATTACGAATTTGTGATACGCCAATTTGTTCTTCTCA 56794
266 alArgValProSerSerSerAlaSerGluAsnAlaIleGlyAspProTyr 282
:|||||:|||||
56793 TTAAAGTTCATCAAGCTCCGCTCAGAGAATGCTATTGGTGATCCATAT 56744
283 LysThrHisIleSerSerAlaAlaLeuSerLeuAsnSerLeuHisThrVa 299
|||||:|||||
56743 AAAACACACATTTCCAGTCAGCGCTGCTGTGAATCCTTGCACTACTGT 56694
299 larGlyTyrSerLysValThrProSerTyrAspAlaThrAlaValThra 316
|||||:|||||
56693 CAGATATTATCTTAAAGTCAACCCCTTCTAGAGCAACTGCTGTAAACAA 56644
316 snGlnAsnIleSerIleHisIleuProSerAlaMetProProValSer.Se 332
|||||:|||||
56643 ATCAGAACATTTCATTCATTACCTTCACGCATGCCCGGCTCTCAAG 56594
332 rPheSerLeuGlyThrAlaLeuLeuGlyTrrPalaArgAlaLeuHisI 349
|||||:|||||
56593 CTCTCCCTCGGCACATGCACTCTGGGGTGGCCAGACGTGCACTACACA 56544
349 leProThrVal..... 352
|||||:|||||
56543 TCCCACTGTCTGTAGATTTCAGGGGTCCAGATGATGACATTGGACCT 56494
352 352
56493 TTCACGCCACGACGAAACCCCTGACGTGATTTGTTGTTGCTACTCT 56444
352 352


```

misc.feature      155874..187508
                  /note="assembly_fragment:01108
                  fragment_chain:2
                  clone_end:SP6
                  vector_side:right"
BASE COUNT      40496 a 54492 c 52718 g 38998 t 804 others
ORIGIN

```

```

alignment_scores:
  Quality: 129.00      Length: 413
  Ratio: 0.701        Gaps: 22
  Percent Similarity: 44.552  Percent Identity: 23.002

```

```

alignment_block:
US-09-389-000-2 x AL158822/rev ..

```

```

Align seg 1/1 to reverse of: AL158822 from: 1 to: 187508

```

```

31 SerLeuPheValPheIleIleProGluAsnPhelYsGlyCysIleSerG1 47
   ::::: ||||| :::::
16454 ACCCTGCTGCTGCTTTCACACCGCAA.....AGCAG 16423
   :::::
47 yHsGlyMetAspIleAlaLeuThrLupProLeuThrMetGluLysMetS 64
   :::::
16422 CCATGGTTGGGCTGGC..... 16404
   :::::
64 eAsnValValLysTyrTrpThrCysProSerAsnThrValLysThr 80
   ||||| :::::
16403 .....TGCACAGCCTGTCGCCGAGGTCACAGGGGCTCAGG 16371
   :::::
81 GluAsnAlaThrGlyProGluLysGlyLeuProLeuGlnArgSerTy 97
   :::::
16370 CCCAGGCTGCTGCTGATGATCGGAGCTCGGGGCT..... 16338
   :::::
97 rSerGluHsLeuGlyTyrPheProThrAspLeuPheAlaCysSerGlu 113
   ||||| :::::
16337 ..TCAGACCAAGGGGGG.....AGCTCTGATCATCTTGTAGCCCC 16298
   :::::
114 ..SerLeuArgAsnGlyAsnGlyLeu.....GluLeuAsn 124
   ||| :::::
16297 TCTGCTGACACAGCCCAAGGCTCTATCTGGCAGGTTCTAGCAGTTG 16248
   :::::
125 AlaSerLeuSerGluPheGluLysAsnLysLys..IleSerLeuLeuHs 141
   ::::: ||||| :::::
16247 TCCCTCTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 16198
   :::::
141 eSerLysGlyLysLeuArgGluArgGluArgGlyLysCysGluGln 157
   :::::
16197 GAACC.....AGGAGCGGATGTCGTTGACCTGTGAGCGT 16163
   :::::
158 LeuArgThrLeuLeuProTyrValLysGlyArgLysAsnAspAlaIle 174
   ::::: ||||| :::::
16162 CTGCGGGGCTGCTGCTGCCCAAGTTCGATGCGCGGAGGAGACATG 16113
   :::::
174 rValLeuGluAlaThrValAspTyrValLysTyrIle..... 186
   ||||| :::::
16112 GGTCTGAGAGATGTCGTGAGTCTCGCGGCTTGCACGCGCCCTGGG 16063
   :::::
187 .....ArgGluLys 189
   :::::
16062 CCAGTCAGAGCAGCAGCTGTGAGTTGTAAGGTAGACCCCTCAGTCC 16013
   :::::
190 IleSerProAlaValMetAlaGlnIleThrGluAlaLeuGlnSerAsn 206
   ::::: ||||| :::::
16012 TCCGCGATCAGGTGATGGACATCTCAC..ACCTTGAGGCA..... 15971
   :::::
206 fArgPheCysLysLysGlnGlnThrProIleGluLeuSerLeuProGly 223
   ||||| :::::
15970 .....GTGTG.....CCCCGTGATCCCCCTGCTTCAG 15940
   :::::
223 hr.....ValMetAlaGlnArgLysAsnSerValMetSerThrTyrSer 237
   ::::: ||||| :::::

```

```

15939 CCTCTGGGCTTCCACGGCTAGACGAGCCAGCTTTGGCGAGGGCACTT 15890
238 ProGluArgGlyLeuGlnPheLeuThrAsnThrCys...TrpAsnGlyC 253
   ||||| :::::
15889 CCTGATTTGTCAGACCTTACTGTTGTTATCTGTTCTCTGGAAGAGAC 15840
   :::::
253 sSerThrProAspAlaGluSerSerLeuAspGluAlaValArgVal... 268
   ||||| :::::
15839 AAGCTTCTCTTCCAGCCCTCTGCGCCCTGATGCTTCTTATGATGAG 15790
   :::::
269 .....ProSerSerAlaSerGluAsnAlaIleGlyAsp 280
   ||||| :::::
15789 GTTCCGAGATGCTCCGTCCCAATTGGGGGAGATGGGCTCTCTGAT 15740
   :::::
281 Pro.....TyrLysThrHisIleSerSerAl 289
   ||||| :::::
15739 CCTCCCTGGCTGAGCAGCCCTAGATCATCTTCTGACGCTCCACAGAG 15690
   :::::
289 AlaLeuSerLeuAsnSerLeuHisThrValArgTyrTyrSerLysVal 306
   ||||| :::::
15689 AGAGGCTCTCTCCCTCTCCAGAGCTGCGGGGCTCTGG.....CTGA 15646
   :::::
306 hrProSerTyrAspAla.....ThrAlaValThrAsnGlnAsnIleSer 320
   ||||| :::::
15645 CCCCAGATGCTGCGGGGCTGAGGACCTGCTGTGGGCAATGAACCTGTG 15596
   :::::
321 IleHisLeuProSerAlaMetProProValSerSerPheSerLeuGlyTh 337
   ::::: ||||| :::::
15595 TCATCTGTCAGGTCAGAGGGTCCACAGCATTTGCAGTAGAGCCCTTGG 15546
   :::::
337 rAlaLeuLeu.....GlyTrpAlaArgArgAlaLeuHisIlePro 350
   ||||| :::::
15545 GCACACCTCTGATTCAGAGATGATGGGCTTTCGGGGGTGAACCTCAG 15496
   :::::
351 ThrValCysAsnSerPheGlyArg..... 358
   ||||| :::::
15495 CTCGTTGTACACTGTACTCTGCTGCCCCCAAGTTCACAGGGGAGCCG 15446
   :::::
359 .....IleLysSerThrCysLeuL 365
   ||||| :::::
15445 TCACAGTACACCTCCAGCCCTGGGGTAGACATCTGTACACCGTGTG 15396
   :::::
365 yAsPhe.....ThrLeuSerThrThrTyr 372
   ||||| :::::
15395 GGTCTCTCAGCTGACTGTCTCTGCTCTTGTGTTTAC 15361
   :::::
seq_name: gb_hhg20:AL354761
seq_documentation_block:
LOCUS      AL354761      172727 bp      DNA      HTG      06-SEP-2000
DEFINITION Homo sapiens chromosome 9 clone RP11-98L5, *** SEQUENCING IN
PROGRESS ***; 10 unordered pieces.
ACCESSION      AL354761
VERSION      AL354761.6 GI:10093788
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 172727)
AUTHORS      Plumb,B.
TITLE      Direct Submission
JOURNAL
SUBMITTED (05-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
REQUESTS: clonerequest@sanger.ac.uk
On Sep 9, 2000 this sequence version replaced gi:19800828.
COMMENT
-----
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
Project Information
Center project name: Ba98L5

```

```

----- Summary Statistics -----
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid: L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 169466 bases at least Q40
Consensus quality: 170480 bases at least Q30
Consensus quality: 171223 bases at least Q20
Insert size: 171827; sum-of-contigs
Insert size: 180951; 1.3% error; agarose-rip
Quality coverage: 8.49x in Q20 bases; sum-of-contigs Quality
coverage: 8.40x in Q20 bases; agarose-rip
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 9669: contig of 9669 bp in length
* 9670 9769: gap of 100 bp
* 9770 24788: contig of 15019 bp in length
* 24789 24888: gap of 100 bp
* 24889 45348: contig of 20460 bp in length
* 45349 45448: gap of 100 bp
* 45449 113889: contig of 68441 bp in length
* 113890 113889: gap of 100 bp
* 113990 127156: contig of 13167 bp in length
* 127157 127256: gap of 100 bp
* 127257 131636: contig of 4380 bp in length
* 131637 131736: gap of 100 bp
* 131737 146937: contig of 15001 bp in length
* 146938 147037: gap of 100 bp
* 147038 153346: contig of 6309 bp in length
* 153347 153446: gap of 100 bp
* 153447 159184: contig of 5738 bp in length
* 159185 159284: gap of 100 bp
* 159285 172727: contig of 13443 bp in length.
*
* Location/Qualifiers
*
* 1.172727
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="9"
* /clone_1fb="RPCT-11.1"
* 1..9669
* /note="assembly-fragment:02116
* fragment_chain:1"
* 9770..24788
* /note="assembly-fragment:01142
* fragment_chain:1"
* 24889..45348
* /note="assembly-fragment:02183
* fragment_chain:1"
* 45449..113889
* /note="assembly-fragment:02357
* fragment_chain:2"
* 113990..127156
* /note="assembly-fragment:01763
* fragment_chain:2"
* 127257..131636
* /note="assembly-fragment:03395
* fragment_chain:2"
* 131737..146937
* /note="assembly-fragment:02045
* fragment_chain:3"
* 147038..153346
* /note="assembly-fragment:01090
* fragment_chain:3"
* 153447..159184
* /note="assembly-fragment:00178
* fragment_chain:3"

```

```

misc_feature      159285..172727
                  /note="assembly_fragment:01839
                  fragment_chain:3
                  clone_end:T7
                  vector_side:right"
BASE COUNT      40126 a 45744 c 46532 g 39420 t 905 others
ORIGIN

alignment_scores:
    Quality: 123.50      Length: 417
    Ratio: 0.660        Gaps: 24
    Percent Similarity: 44.844      Percent Identity: 23.261

alignment_block:
US-09-389-000-2 x AL354761/rev ..

Align seg 1/1 to reverse of: AL354761 from: 1 to: 172727

31 SerLeuPheValPheIleIleProGluAsnPhenylsGlyCysIleSerI 47
   :||| ||||| |||||
98372 ACCCTCGCTGCTCTTTCCACACGCCAA.....AGCAG 98341
   :||| ||||| |||||
47 YHisGlyMetAspIleAlaLeuThrGluProLeuIleMetGluYMetS 64
   :||| ||||| |||||
98340 CCATGGTTGGGCGCTGGG..... 98322
   :||| ||||| |||||
64 eAsnValValIleTyrTrpThrCysProSerAsnThrValIleThr 80
   :||| ||||| |||||
98321 .....TGACAGCGCTGCGCGCAGGTCACAGGGTCAGG 98289
   :||| ||||| |||||
81 GluAsnAlaThrnglyProGluGluLeuGlyLeuProLeuGlnArgSer 97
   :||| ||||| |||||
98288 CCCAGCGCTGCTCGATGGATCGCTCGGCT..... 98256
   :||| ||||| |||||
97 rSerGluHisLeuGlyTyrPheProThrAspLeuPheAlaCysSerGlu 113
   :||| ||||| |||||
98255 .TCAGACCACAGGGGGC.....AGCTTGAGTCATCTGTACCCCC 98216
   :||| ||||| |||||
114 ..SerLeuArgAsnGlyAsnGlyLeu.....GluLeuAsn 124
   :||| ||||| |||||
98215 TCTCCCTGCACAGCGCCACAGGCTCATCTGCAGAGTCTAGACGTTG 98166
   :||| ||||| |||||
125 AlaSerLeuSerGluPheGluGlyAsnIleCysIleSerLeuHis 141
   :||| ||||| |||||
98165 TCCCTTCTCGGGGAGGCGACAGGCGCCAGAAATTTGTTCTTGACCG 98116
   :||| ||||| |||||
141 eSerTyrGluIleuArgArgGluArgIleTyrCysCysGluGln 157
   :||| ||||| |||||
98115 GAAC.....AGGAAGCGGATGCTTGAGCTGTGAGCGT 98081
   :||| ||||| |||||
158 LeuArgThrLeuLeuProTyrValIleGlyArgIleCysAsnAlaAla 174
   :||| ||||| |||||
98080 CTGGGGCGCCGCGGCCACAGTGTGAGGCGCGGGGAGAGACAGCGCTC 98031
   :||| ||||| |||||
174 rValLeuGluAlaThrValAspTyrValIleTyrIle..... 186
   :||| ||||| |||||
98030 GGTCTGAGATGTCTGTGAGTTCGCGCTTGCCAGCGCCCTGGGCG 97981
   :||| ||||| |||||
187 .....ArgGluIle 189
   :||| ||||| |||||
97980 CCACTCAGAGCACACAGCTGTGAGTTGTAAGGTAGACGCCGTCAGTCC 97931
   :||| ||||| |||||
190 IleSerProIleValMetIleGlnIleThrGluAlaGlnIleSerAsn 206
   :||| ||||| |||||
97930 TCCGGCTCACGGTGGAGGACATCTACC...ACCTTGCGGGA..... 97889
   :||| ||||| |||||
206 tArgPheCysIleGlyGlnIleThrProIleGluLeuSerLeuProGly 223
   :||| ||||| |||||
97888 ....GTGTGT.....CCCGTGATGCCCGCTCTCCTCAG 97858
   :||| ||||| |||||
223 hr.....ValMetAlaGlnArgGluAsnSerValMetSerThrTyrSer 237
   :||| ||||| |||||

```


629 CTCGCCGGGCTGGCGCGCCCGCCCTCTGTAGCGCCGTCCTGCT 678

293LeuansSerLeuHisThrValArgTyrSerLysVal 305
||| ::| |||::: ::| |||
679 CTTGGCGCTGCTGCGCGCCAGCCAGTCGTGACCCCTCAGT 722

306 ThrProSerTyrAspAlaThrAlaValThrAsnGlnsnle 319
::| ||| ||| ||| ||| ::| ||| ::|
723 TCTCTTCACTTGAAGCCGACCTCTCCACAGGCGCCACTGCTGTCCATTC 772

320SerHisLeuProSerAla.... 326
::| ||| ||| ||| |||
773 TGGAGCTCCAGCTGCTGCTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 822

327MetProValSerSerPheSerLeuGlyThrAlaLeuGly 341
::| ||| ::| ||| ::| |||
823 AGGCGCTGCGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 872

342 TTPAlaArgArgAlaLeu...His.....IlePr 350
::| ||| ||| ||| ||| |||
873 GGAGGAGGCGCCGACGCTTCCACATCCGAGCCCTGCTGCTGCTGCTGCTGCT 922

350 oThrValGly.....AsnSerPheGlyArgIleSerThr 363
||| ||| ||| ||| ||| |||
923 CAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 972

363 ySLeulys 365
|| |||
973 GCTGCAA 980

seq_name: gp_pr3:AF070992

seq_documentation_block:
LOCUS AF070992 2313 bp mRNA PRI 10-SEP-1999
DEFINITION Homo sapiens HPV-16 E2 binding protein (E2BP-1) mRNA, complete cds
ACCESSION AF070992
VERSION AF070992.1 GI:5852278
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2313)
Zheng, P.-S. and Pater, A.
Direct Submission
Submitted (08-JUN-1998) Division of Basic Science, Faculty of
Medicine, Memorial University of Newfoundland, St. John's,
Newfoundland A1B 3X6, Canada
Location/Qualifiers
1..2313
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..2313
/gene="E2BP-1"
76..1434
/gene="E2BP-1"
/codon_start=1
/product="HPV-16 E2 binding protein"
/protein_id="A053986.1"
/db_xref="GI:5852279"
/translation="MTEVEYQQLHIICSHNEAADGELTRLSNLSALLAAGPGAGAG
GFAAGGOGGAGAPVPLVPCPALADAPICGIDIDLELMILSEGALEKTSGGGCG
PARADGAKGAGAAAAGDADPEAKAPKAVRYLRDRNSIPAEPPAPRGDEP
EGGALNNLVTLIRHPSLMVPLQDQKCKTALVKNKKAATTTALQFYPLPTNAC
TSGNSLSQITSSNSCSYVLEAAKHQDGLPRAFSCTQDIESTKQTLGRKYLPR
QVNLKVGSAALCKQALKRNRSMKQDLTVNVRALGETIQVNGEATATQAGWQSS
QANLGEQASQPGGSRQRNRHNRHMERDRRRRIICDELNLVPPFNATIDATTII
QMTAFATKYYIQRHGDLSLKEFESEVFCGTGRRLKLPDLSLVCPAGSLQSSPME
IK"
124
/gene="E2BP-1"
/note="alternate position for initiation methionine"

BASE COUNT 608 a 558 c 609 g 538 t

```

ORIGIN
alignment_scores:
    Quality: 119.00      Length: 101
    Ratio: 1.776        Gaps: 3
Percent Similarity: 66.337   Percent Identity: 28.713

alignment_block:
US-09-389-000-2 x AF070992 ..

Align seg 1/1 to: AF070992 from: 1 to: 2313

140 HISSerSerySGluSLysLeuAArgGUArgUleLysTYrCYscYsgl 156
||||| ||||| ||||| ||||| ||||| |||||
1144 CATACCGCATGGAAGAGATGAAAGGCACAGATCCGATTGGCTGTGA 1193

156 uGlnLeuAArgThrLeuLeuProTYrValYslGLyAlgLYAsnASPala 173
||||| ||||| ||||| ||||| ||||| |||||
1194 TGGTTGAAATCTCTTAGTGCCGTCGCAATGCC...GAGACTGACAAG 1240

173 laSerValLeugInuaLaThrVaLaspTYrValIstYrILearGluLyS 189
||||| ||||| ||||| ||||| ||||| |||||
1241 CCACACACTGCGAGTGGACCACAGCATTCTCGAAATACATCCAGGAAAA 1290

190 lIeSerPrOALaYaMeLaIagInIleRhgInuaLeuGInSerASnMe 206
||| ||||| ||||| ||||| ||||| |||||
1291 CANGGAGATTCTCTTAAGAAGCAATTGAGCGCTR..... 1326

206 tarGPeCySLysGLySGInGIInThPrOIleGlueSerLeProGLYT 223
||||| ||||| ||||| ||||| ||||| |||||
1327 ....TTTTCCGGTAAACATCGCCGGAAGGCTAAAGCTGACACACCGGACT 1372

223 hr...ValMeLaIagInARgGUAnSerValMetSerThryrSerPro 238
:: ||| ::||| ::||| ::||| ::||| ::|||
1373 CCTGTGTGACCTGTCTCGCACAGGGAGTTTACAGAGACGCCCTCGATG 1422

239 Glu 239
|||
1423 GAG 1425

seq_name: gb_prl:AB012124

seq_documentation_block:
LOCUS AB012124 2316 bp mRNA PRI 09-JAN-1999
DEFINITION Homo sapiens TCFL5 mRNA for transcription factor-like 5, complete cds.
ACCESSION AB012124
VERSION AB012124.1 GI:4126408
KEYWORDS transcription factor-like 5; TCFL5.
SOURCE Homo sapiens cdna to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
AUTHORS Nakamura,Y., Nishimori,H. and Maruyama,O.
TITLE 1 (bases 1 to 2316)
JOURNAL Direct Submission
Submitted (12-MAR-1998) to the DDBJ/EMBL/Genbank databases, Yusuke
Nakamura, Institute of Medical Science, The University of Tokyo,
Laboratory of Molecular Medicine, Human Genome Center; 4-6-1
Shirokanedai, Minato-Ku, Tokyo 108-8639, Japan
(E-mail:yusuke@ims.u-tokyo.ac.jp, Tel:81-3-5449-5372,
Fax:81-3-5449-5433)
2 (sites)
REFERENCE
AUTHORS Maruyama,O., Katagiri,T., Nishimori,H., Miki,Y., Ueno,A. and
Nakamura,Y.
TITLE Cloning, mapping, and genomic organization of THLH1, a novel human
testis-specific gene containing a basic Helix-Loop-Helix motif
JOURNAL Unpublished (1998)
COMMENT Sequence updated (27-Jun-1998).
FEATURES
Location/Qualifiers
1..2316
/organism="Homo sapiens"

```


gene /db_xref="taxon:9606"
 99..1457
 /gene="TCFL5"
 CDS 99..1457
 /gene="TCFL5"
 /codon_start=1
 /product="transcription factor-like 5"
 /protein_id="BAA36557.1"
 /db_xref="GI:4126409"
 /translation="MTEVEYTOQHILCSHMEAAADGELEMRNSALLAAAGPGAG
 GRAAGGGAAPYPIVPCSPALADAPCIGHIDFOLRMMLSEAGAEKTSGGGGA
 RARADNAKEGAAAAAAGPDAPARAPARVLEDRFNSIPAPAPGPPPP
 EREGALNNLVTLRHPSELMNVPLQONKCTALVKNKTAITTAQFTPLFTNACS
 TSGSNLVSQIYSSNSCSVLEAKHODIGPRAFSFCYOQIESTKOTUGSRKVLPE
 QWIKVGEALCKQALKRNSRMRLDITVERALGIEIOWGEGATATGAMQSSSS
 QANLGEAQSPPGGRSRRRERHMRERRRIRICDELNLVFCVAETDKATTL
 OMTTALFKYIOERHSDSLKKEPESVCGTGRKLRLRDSLVTCAPSLSQSSPME
 IK"

BASE COUNT 591 a 564 c 623 g 538 t
 ORIGIN

alignment_scores:
 Quality: 119.00 Length: 101
 Ratio: 1.776 Gaps: 3
 Percent Similarity: 66.337 Percent Identity: 28.713

alignment_block:
 US-09-389-000-2 x AB012124 ..

Align seg 1/1 to: AB012124 from: 1 to: 2316

```

140 HisSerSerLysGluLysLeuArgArgGluArgGlyLeuLysTyrCysCysG1 156
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1167 CATACCAAGATGAAAGAGATAGAGGCGCAGAACCCGATTCGTGTGA 1216
156 uGlnLeuArgThrLeuLeuProTyrValLysGlyArgLysAsnAspAlaA 173
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1217 TGAAGTGAATCTCTAGTACGCGTCTGCATGCC...GAGACTGACAAGG 1263
173 LasSerValLeuGluAlaThrValAspTyrValLysTyrIleArgGluLys 189
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1264 CCACACTCTCAGTGACGACCAAGCATTCGAAATACATCAGGAAAGA 1313
190 IleSerProAlaValMetAlaGlnIleThrGluAlaLeuGlnSerAsnMe 206
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1314 CATGAGATTCCTTAAAAAGAAATTGAGAGCGTA..... 1349
206 TarGpHeCysLysLysGlnGlnThrProIleGluLeuSerLeuProGlyT 223
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1350 ....TTTGGCGGTAAACCTGCGCCGAGGCTAAAGCTGACACAGACGGACT 1395
223 hr...ValMetAlaGlnArgLysAsnSerValMetSerThrTyrSerPro 238
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1396 CTTGGTGACCTGCTCTCCACAGGGGAGTTTACAGAGCAGCCCTCGATG 1445
239 Gln 239
|||
1446 GAG 1448
```

seq_name: gb_to:MUSMYN

seq_documentation_block:
 LOCUS MUSMYN 1941 bp mRNA ROD 27-APR-1993
 DEFINITION Mus musculus B/HLH/Z protein (Myn) mRNA, complete cds.
 ACCESSION M63903
 VERSION M63903.1 GI:1399986
 KEYWORDS B/HLH/Z protein.
 SOURCE Mouse, cDNA to mRNA.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 1941)

AUTHORS Ziff, E.B., Lave, D. and Prendergast, G.C.
 TITLE Association of myn, the murine homolog of max, with c-myc
 stimulates methylation-sensitive DNA binding and ras
 contrainsformation
 JOURNAL Cell 65, 395-407 (1991)
 MEDLINE 91208676
 FEATURES
 source location/Qualifiers
 1..1941
 /organism="Mus musculus"
 /strain="Bals/c"
 /sub_species="domesticus"
 /db_xref="taxon:10090"
 /cell_line="A31"
 /cell_type="Fibroblast"
 /tissue_lib="lambda gt10 made with RNA from
 serum/cycloheximide-stimulated 3T3 cells"
 1..1941
 /gene="Myn"
 /note="putative"
 1..1941
 /gene="Myn"
 161..643
 /gene="Myn"
 /note="putative"
 /product="B/HLH/Z protein"
 /protein_id="Aa39797.1"
 /db_xref="GI:199987"
 /translation="MSDNDIDIEVSDDEQARFOSADKRRAHNALEKRRDHIKDSFH
 SLRDSYPSIQEKASRAQITDKATEYIQNRKRNDPDODIDIDIKRNLLEQVAF
 EKARSSAOLQTNIPSSDNLVTNAKGITISAFDGGSSSESEPEPQSRKILRMAS
 "

BASE COUNT 464 a 495 c 468 g 514 t
 ORIGIN

polyA_signal 1917..1922
 /gene="Myn"
 /note="putative"

alignment_scores:
 Quality: 118.00 Length: 310
 Ratio: 0.756 Gaps: 13
 Percent Similarity: 50.323 Percent Identity: 21.935

alignment_block:
 US-09-389-000-2 x MUSMYN ..

Align seg 1/1 to: MUSMYN from: 1 to: 1941

```

140 HisSerSerLysGluLysLeuArgArgGluArgGlyLeuLysTyrCysCysG1 156
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
242 CATATGACACTGAAACGAAAGTAGGAGCACATCAACACACTTCA 291
156 uGlnLeuArgThrLeuLeuProTyrValLysGlyArgLysAsnAspAlaA 173
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
292 CAGTTTCGGGACATCAGTCCATCCTCAAGAGAGAGGACATCCCGG 341
173 LasSerValLeuGluAlaThrValAspTyrValLysTyrIleArgGluLys 189
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
342 CCCAATCTTAGACAAAGACACAGATATCATCATATCCGAAAGAAA 391
190 IleSerProAlaValMetAlaGlnIleThrGluAlaLeuGlnSerAsnMe 206
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
392 .....AACGATACGCCACACGAAAGACAT 414
206 TarGpHeCysLysLysGlnGlnThrProIleGluLeuSerLeuProGlyT 223
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
415 TGATGACCTCAAGCGGCAAGATGCTCTCTGACACAAACAACTCCGTGAC 464
223 hrValMetAlaGlnArgLysAsnSerValMetSerThrTyrSerProGlu 239
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
465 TGAAGAAGCA...AGATCAAGTCCCAACTGACAGACCAACTACCCCTCC 511
```


137 *erleuEnhHisSerIysGlyLysLeuArgAlaArgLysIyr* 153
504 AACCCCGAGGCCAACCGGTGAAGCGGTGAACAC 542
154 *CysCysGluGlnLeuArgThrLeuLeuPro*.....Ty 164
543 AACCTTCGCCCGCTCTGCCAGACAACTTCCCACTGCATCTCGTGACT 592
164 *rValIysGly*.....ArgLysAsnAspAlaAlaSer 175
593 GACCAAGGAGGTGGCCGGGGCCGCACAGAAGAATCTCCAGGTGGACA 642
175 *alLeuGlnAlaThrValAspTyrValLysTyrIleArgGluLysIle*... 190
643 CACTCCGCAATTCCTCGGATATCATCTCGCGGCTCCAGAACTCGTGCAC 692
191SerProAlaValMetAlaGlnIleThrGlnAlaIle 202
693 GACCTCAATGGGGGTGCGCCGCCCACTTCACATGCCCGCCGCAATGCCGT 742
202 *uGlnSerAsnMetArgPheCysLysLysGlnIleThrProIleGluLeu* 219
743 CGTCAGCCAGCTGCAGCTCTGCTGGAGAGGAGAAC..... 779
219 *erLeuProGlyThrValMetAlaGlnIleArgGluAsnSerValMetSerThr* 235
780TCGACGAAACGCTCCAGCAGCTCC 803
236 *TyrSerProGluArgGlyLeuGlnPheLeu*..... 245
804 GGCACAGCAGCAGCATAGCAGCCACCACTCTACTACAGAGCTCCTCGC 853
246ThrAsnThrCysTrpAsnGlyCysSerThrProA 257
854 CGCAGCCCGGACAGACAGCAGCAGCATCAGCGCGGACAGACTACACAG 903
257 *spAlaGlnSerSerLeuAspGlnAlaValAlaArgValProSerSerSerAla* 273
904 ACGAGCAGCTCC.....AACTCCAACTCC 926
274 *SerGluAsnAlaIleGlyAsp*.....Pr 281
927 AACCTCAACTCCTCTCGACAGAAACAGCAGCTACAGCCGCAAGATTTC 976
281 *oTyrLysThrHisLeuSerSerAlaAlaLeuSerLeuAsnSerLeu*... 296
977 CCATCAGCAACACACTAGCAGCTCATCTCGTGAACAGCAACGATCGGCG 1028
297HisThrValArg 300
1027 TGGGTGACGTGGCTGTGTGTGGCAGCGTTCGTGGCGAGCGAGCGCC 1078
301 *TyrTyrSerLysValThrProSerTyrAspAlaThrAlaValThrAsn* 316
1077 TGGCAGTCGCCACACTCTCTCTTCAACTCGACATGTCTTCGACTCGGG 1122
317GlnAsnIleSerIle...HisLeu..... 323
1127 CACTATGAGCGCGCGCCGACAGCACTGTCTGCGCACCCACATGGAGCTG 1176
324ProSerAlaMetProPro...ValSerSerPhe 334
1177 GACATGGAACCTGAGCCGGAGGCGATGGCCGCCCACTCAAGTCAAGTTC 1228
334 *ribEuglyThrAlaLeuLeuGlyTyrAlaArgArgAlaLeuHisIleProT* 351
1227 GAGCCCTCAGACACTTCCAGCTGGACGAGGAGGACTGCACAC...CCGA 1273
351 *hrValCysAsnSer*..... 355
1274 CGACGAGGAGATCTCGACTACATCTCATGTCGAGAGGACAGTATCAT 1322
356PheGlyArgIleLysSerThrCysLeuLysPheThrLeu 369

```

|||||: :||: ||::|||: |::|: ||
1324 CGCATCCGTTCAATCTGTCCCCCCTTTCGCATCGTGATCTGACATCTAG 1373
    369 r 369
      |
1374 T 1374

seq_name= gb_ov:XELXMAX4A

seq_documentation_block:
LOCUS XELXMAX4A 878 bp mRNA VRT 06-APR-1994
DEFINITION Xenopus laevis Xmax4 mRNA, complete cds.
ACCESSION L09739
VERSION L09739.1 GI:214912
KEYWORDS Myc protein; variant.
SOURCE Xenopus laevis (library: Lambda gtl1 (by C. Kintner)) embryo
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesopoda; Pipidae; Fipidae;
Xenopodinae; Xenopus.
REFERENCE Tolissen,K.F. and Krieg,P.A. Analysis of a variant Max sequence expressed in Xenopus laevis
AUTHORS Oncogene 9 (1), 33-38 (1994)
JOURNAL J94I34435
MEDLINE Location/Oualifiers
FEATURES
source 1..878 /organism=Xenopus laevis"/db_xref=taxon:8355"
5'UTR 1..110
CDS 111..602 /standard_name="XMax4" /codon_start=1 /function=forms heterodimers with the Myc protein family"/protein_id="AA17425.1"/db_xref="GI:214913" /translation=WSDNDIDLEVESDDSRFPYSADKRAHHNNALEKKRDHDKDSTH/GLRSVPLQGEKASRAQLLDKATEYIQVMKNKHQQDIIDLKRONALLDEOVQIS/NPKPNSGAPEOKNKVNLIOLSPSKVALEAKSSQSLOSNESSSEETEPQRKLMM/DAS"
BASE COUNT 252 a 228 c 163 g 235 t
ORIGIN
alignment_scores:
Quality: 116.50 Length: 302
Ratio: 0.793 Gaps: 13
Percent Similarity: 48.675 Percent Identity: 21.834
alignment_block:
US-09-389-000-2 x XELXMAX4A ..
Align seg 1/1 to: XELXMAX4A from: 1 to: 878
86 ProGLUGLuleuglYLeufProleunglnarGerTYSerGlunHisDeuGl 102
|||||: |||||
66 CCCGACACAGCACGCCCTCCC..... 86
102 yTyTPhePrOThraSpLeuPhenaLaCySeerGuUseTerLeuaArganGLyA 119
||||| ::|||
87 .TCCTCCCA..... TCATTATAGGAMAACATGACGATMACG 123
119 sngLYeuclueLuensaNa..... SerLeusercgluhegLUlysasn 133
:::||:|||||: ||::|||
124 ATGACTCATGGAGGGAAGACAAGAAGACTCGCAAGATTCCCGATTGCG 173
134 LySLysIlleserLeuauHisserSetrLySGluYLysLeuarARGGLuar 150
|||||: ||::|||: ||::|||
174 GCAGAACAAACGACCCTACATAATGACACGAGCGCAAACGACGAGACCA 223
150 gltelytyrcyscyslgnlneuatrgThrLeuLenPuotyVallysg 167
```

```

224 TATCAAGGACGCTTCCAGCGCTTAAGGACGCTCCGCTACCATCTCCAG 273
167 LyrArgLysAsnAspAlaLeuSerValLeuGluAlaThrValAspTyrVal 183
274 GCGAAAAGGCGATGAGGCGCCCAATACTGGACAAAGCTACAGAGTACATT 323
184 LysTyrIleArgLysIleSerProAlaValMetAlaGlnIleThrG1 200
324 CAGTACATGAGGAGGAAA...AATCACACACACAGCAATATCGATGA 370
200 uAlaLeuGlnSerAsnMetArgPheCysLysLysGlnGlnThrProIleG 217
371 CTTAAAGCGACAGAAATGCTTATTA.....GAACACAAAGTACAAATCT 414
217 LuleuSerLeuProGlyThrValMetAlaGlnArgLysValMet 233
415 CCATCCAAAACCAACCCAGACGAGCCCTGAGCGAGAAAATGTTTG 464
234 SerThrTyrSerProGlu.....ArgGlyLeuGlnPheThrAsnTh 248
465 CAGTTGCTCAGCCCTTCAAAAGTTCTGCTGACTGAA..... 500
248 rCysTyrPasnGlyCysSerThrProAspAlaGluSerSerLeuAspGlu 265
501 .....AAGCAAAATCCAGCTCTCAGCTGAGTCCACTAC..... 536
265 lAlaValArgValProSerSerSerAlaSerLysAlaIleGlyAspPro 281
537 .....TCAGCTCCGAGTCCAGAACTGAAGAACCC 566
282 TyrLysThrHisLeuSerSerAlaAlaLeuSerLeuAsnSerLeuHisTh 298
567 CAA.....ACGAGAAAGAACTACGATGACCG..... 596
298 rValArgTyrTyrSerLysValThrProSerTyrAspAlaThrAlaVal 315
597 .....ACGTAGACCTGTCGACGACATTAACACAGACGCTGT 633
315 hAsnGlnAsnIleSerIleHisLeuProSerAlaMetProProValSer 331
634 CCAGTTTAACTCCGGCCCATATTAACCCCTTTGTTAGCCCGCTCAGC 683
332 SerPheSer.....LeuGlyTh 337
684 TGTTTACAGACTTTTGTGTTTGTGTTTGTGTTGCTCAGCCCAAT 733
337 rAlaLeuLeuGlyTyrAlaArg.....ArgAlaLeu 348
734 ATCGTTATGCTCTTGAGCGGTTTTTTTTTAATGAATTTGCGCTTC 783
348 lAsIleProThrValCysAsnSerPheGlyArgIleLysSerThrCysLeu 364
784 AC.....TGTCTC 791
365 LysPhe 366
792 CGATTTC 797
seq_name: gb_to:MUSHSP86A
seq_documentation_block:
LOCUS MUSHSP86A 2975 bp mRNA ROD 10-NOV-1994
DEFINITION Mouse heat shock protein 86 mRNA, complete cds, and 28S ribosomal
RNA, partial sequence.
ACCESSION J04633
VERSION J04633.1 GI:194030
KEYWORDS heat shock protein.
SOURCE Mouse (strain BALB/c) CDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 2975)

```

```

AUTHORS Moore,S.K., Kozak,C., Robinson,E.A., Ullrich,S.J. and Appella,E.
TITLE Chromosome assignments, and evolutionary origins
JOURNAL J. Biol. Chem. 264 (10), 5343-5351 (1989)
MEDLINE 89174568
REFERENCE 2 (bases 1 to 2975)
AUTHORS Moore,S.K.
TITLE Direct Submission
JOURNAL Submitted (11-APR-1989) Stephen K. Moore, Laboratory of Cell
Biology, National Cancer Institute, NIH, Bethesda, MD 20892, USA
FEATURES
source
1..2975
/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/haplotype="H-2d"
/map="1 bp upstream of EcoRI site"
/cell_line="fibrosarcoma cell line Meth A"
<1..1588
/note="heat shock protein 86 mRNA"
<1..>229
/product="28S ribosomal RNA"
261..2462
/note="heat shock protein 86"
/codon_start=1
/protein_id="AA53068.1"
/db_xref="gi:309318"
/translation="MPEETOTQDQPMEEVEVFARQAEIQAQMSLIITFTYSKEIF
LRELINSSDALRKIRYESLTPSKIDSGKELHNLIPSKODRLTIVDGIKTKAD
LINLIGTIAKSGTAKAFEMLOAGDIDIMIGFSAVLAVKEVIVIRKHNDDQY
AMESAGSGFTVRDITGEPMRGRKVLHLKEDQTELEERRIKEIKKRSOPGYPI
TLPEKRDKEVSDDEAEKEEKEEKEEKEEKSDDKEIEDYGSDEEKEKDGOK
KKKIKRKTYIDQELNTRPIWTRNPDITNEEGEPYKSLTDWMEHLAVKRPVSG
QLEKRALFVPRRAPDLFENRKNKIKLIVRFVIMDCEELIPYLNFGVDS
EDLPNLISREMLQOSKILKVIKRLVKCLEFTELEADKENYKKEFQPSKIKGLGI
HESONRKKLSSELRIRYTSASGDEMSLKYCTRKMKGNKRIYITGETDKOYNSAF
VERLRKGLFVIVWIEPIDEYVOOLKEFEGKTLVSYKGELEPDEEKKKQOEKK
TKFENLCKTKIMDKLEKKEKVVVSNRLVTSPPCIIVSTGYWTANMERIMAAOLRDS
TMGMMAKKHLEINPDHSITETLROKREAKRNKSKYDVIILLYETRLSSGSLEDP
QTHANRIRYRKIKGLGIDEDDPTVDDTSAAVTEMPLEBDDDTSRMEVD"
BASE COUNT 959 a 549 c 719 g 748 t
ORIGIN
alignment_scores:
Quality: 115.50 Length: 361
Ratio: 0.628 Gaps: 22
Percent Similarity: 48.294 Percent Identity: 22.835
alignment_block:
US-09-389-000-2 x MUSHSP86A ..
Align seg 1/1 to: MUSHSP86A from: 1 to: 2975.
6 ValProSerSerLeuSerAlaGluLeuGluAlaIleLysLeuIleAr 22
1443 CTCCTCTTAATATTTCCCGTAAATGCTGCACAAAGTAAATCTGAA 1492
22 gpPheGlyLysLys.....LysAsnThrHisSerLeuPheValPheI 36
1493 AGTTATCAGAAAGAAATTGGTCAGAAAGTCTAGACTATTACTGAAC 1542
36 lAlaLeu.....ProLysnPhelysGlyIleSerGlyHisLysMet 50
1543 TGGCAGACAGATTAAGAGAACTCAAAAAGTTTATGACGACAGTTCCAA 1592
51 AspIleAlaLeu.....ThrGluProLeuThrMetGluLysMetSe 64
1593 AATATTAAGCTTGGAAATTCACAGGAGCTCTCAGAAATCGGAAGAGCTTTC 1642
64 rAsnValValLysTyrTyrThrCysProSerAsnThrValLysThrG 81
1643 AGAGCTGTGGCGGTACTACACA..... 1664

```

[illegible]

```

seq_name: gb_pr3:AF040963

seq documentation block:
LOCUS      AF040963          879 bp      mRNA          PRI          20-JAN-1998
DEFINITION Homo sapiens Mad4 homolog (Mad4) mRNA, complete cds.
ACCESSION  AF040963
VERSION    AF040963.1
KEYWORDS   GI:2792361

SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  1 (bases 1 to 879)
AUTHORS   Pribill,I., Barnes,G.T., Chen,J., Church,D., Buckler,A.,
            Bendale,S., Bates,G.P., Lehrach H., Gusella,M.J., Duyao,M.P.,
            Ambrose,C.M., MacDonald,M.E. and Gusella,J.F.
TITLE     Comparison of Exon Trapping and Sequence-based Methods of Gene
            Finding
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 879)
AUTHORS   Pribill,I., Barnes,G.T., Chen,J., Church,D., Buckler,A.,
            Bendale,S., Bates,G.P., Lehrach H., Gusella,M.J., Duyao,M.P.,
            Ambrose,C.M., MacDonald,M.E. and Gusella,J.F.
TITLE     Direct Submission
JOURNAL    Submitted (05-JAN-1998) Molecular Neurogenetics Unit, Massachusets
            General Hospital, 13th street, Charlestown, MA 02129

FEATURES
            source
                1..879
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="4"
                /map="4p16.3"
                1..879
                /gene="Mad4"
                14..643
                /gene="Mad4"
                /note="T72"
                /codon_start=1
                /product="Mad4 homolog"
                /protein_id="AAB97009.1"
                /db_xref="GI:2792362"
                /translation="MEINSILITLEAEVLEERDRERAEHGVA5VLPEDGPARREKTKA
            AGLYRKPNNRSSHNELEKRRRLATRLYLDTOLVIGDPSFHTLISLKRKVIH
            KKEQDPRRAISKEQLQERFLKRLBQLSVQSVRVRTRDTSGSAVSTDSDSEQVD
            IEGEFEPGEIDSVSGSSDDADHDHYSLSGTGDSGFQPHCRIGRPALS"

BASE COUNT      166 a          296 c          285 g          132 t
ORIGIN
alignment_scores:
            Quality: 114.50          Length: 259
            Ratio: 0.867          Gaps: 12
Percent Similarity: 50.965          Percent Identity: 23.938

alignment block:
US-09-389-000-2 x AF040963 ..

Align seg 1/1 to: AF040963 from: 1 to: 879

131  GULYtAsnlystlyslEsErLeu...:::139
||||:||||:
131  GAGAAACAAGAAGCGCGCGCTGCGCCAGACGCCGACACAGACGc 180
||||:||||:||||:
140  ...HisSerSetLysGLuLyLeuArGArGLuAgtIeLyStYrCYcG 155
||||:||||:||||:||||:
181  TTCACCAACAGAGCTTGAAAGACAGACAGACCAACACAGCGTGTTACC 230
||||:||||:||||:||||:
155  ysgLulnLeuArGtHrLeuLeuPrtYr...ValYsgLyArGLYtAsn 170
||||:||||:||||:||||:
231  TTGAGCAGCTCAAGCAACTGTGCGCCCTGGGCCGACAGACACCGCCAC 280
||||:||||:||||:||||:
171  AspAlaIseValleuGlulAthrValAspTYrValYlyTYrTleAr 187
||||:||||:||||:||||:

```



```

212 nglnthProIIeGluLeuSerLeuProGlyThrValMetAlaGlnArg 229
    ::::: ||| |||::: |||
539 CTAAGCTCAGTGAAGGCGCTTGACCCGTGAGACCCACGCGTGGCC 588
    ||| ::::: ||| ::::: |||
229 LuSnSerValMetSerThrTyrSerProGluArgGlyLeuGlnPheLeu 245
    ||| ::::: ||| ::::: |||
589 TTGCTCTCATCTGACGACCTGTGCTCCGCGG..... 623
    ||| ::::: ||| ::::: |||
246 ThrAsnThrCysTrpAsnGlyCysSerThrProAspAlaGluSerSerLe 262
    ::::: ||| ::::: |||
623 ..... 623
262 uAspGluAlaValArgValProSerSerSerAlaSerGluAsnAlaIleG 279
    |||::: |||::: |||::: |||
624 ....GAGGACGACGATGACATCCTCCATGCGCCACACCATCACCCCT 669
    |||::: |||::: |||::: |||
279 LysAspProTyrLys.....ThrHisIleSerSerAlaAla 290
    |||::: |||::: |||::: |||
670 TGCACCCCTCACCTGAGGAGCTGCTTCCACATCTCCCCACAGCCCTG 719
    |||::: |||::: |||::: |||
291 LeuSerLeuAsnSerLeuHisThrValArgTyrTyrSerLysValThrPr 307
    |||::: |||::: |||::: |||
720 CTCACGCCCATGACTGCTCCACACA.....TCAGAGTCACACCC 757
    |||::: |||::: |||::: |||
307 oSerTyrAspAlaThrAlaValThrAsnGlnAsnIleSerIleHisLeuP 324
    |||::: |||::: |||::: |||
758 ATGTCGCTTCACCATCT..... 776
    |||::: |||::: |||::: |||
324 roSerAlaMetProValSerSerPheSerLeuGlyThrAlaLeuLeu 340
    |||::: |||::: |||::: |||
777 ..TCAGAGGCGCTTCTGCTCATGCGCTGCTCTCTCCACAGCAGACGTTT 824
    |||::: |||::: |||::: |||
341 GlyTPrAlaArgArgAlaLeuHisIleProThr 351
    |||::: |||::: |||::: |||
825 GCCCATGACAGATCCGCTCTTCGATGCCATCA 857
    |||::: |||::: |||::: |||

```

seq_name: gb_ro:MMU271867

seq_documentation_block:

LOCUS MMU271867 2539 bp mRNA ROD 11-JUL-2000

DEFINITION Mus musculus mRNA for basic-helix-loop-helix protein (Hey2 gene).

ACCESSION AJ271867.1 GI:6900326

VERSION basic-helix-loop-helix protein; Hey2 gene.

KEYWORDS house mouse.

SOURCE Mus musculus

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2539)

Leimaster,C., Externbrink,A., Klamt,B. and Gessler,M.

Hey genes: a novel subfamily of hairy- and Enhancer of split

related genes specifically expressed during mouse embryogenesis

Mech. Dev. 85 (1-2), 173-177 (1999)

JOURNAL MEDLINE 99345954

2 (bases 1 to 2539)

Steldi,C., Leimaster,C., Klamt,B., Maier,M., Nanda,I., Dixon,M.,

Clarke,R., Schmid,M. and Gessler,M.

Characterization of the human and mouse HEY1, HEY2, and HEY3 genes:

cloning, mapping, and mutation screening of a new bHLH gene family

Genomics 66 (2), 195-203 (2000)

JOURNAL MEDLINE 20318621

3 (bases 1 to 2539)

Direct Submision

Submitted (31-JAN-2000) Gessler M., Theodor-Borelli-Institut fuer

Biowissenschaften (BioCenter), Physiological Chemistry I, Am

Hubland, 97074 Wuertzburg, GERMANY

location/Qualifiers

1..2539

/organism="Mus musculus"

/db_xref="taxon:10090"

/chromosome="10"

/clone_lib="129/ola mouse cosmid"

```

gene /clone="MPMGc121M23647Q2"
136..1155
/gene="Hey2"
136..1155
CDS /gene="Hey2"
/codon_start=1
/product="basic-helix-loop-helix protein"
/protein_id="CAB71346.1"
/db_xref="GI:6900327"
/translation="MKRPCEETTSSESDLETDIDVGSENNYPGHATSYMRSNSPTTTS
OIMARKRRGRIIEKRRDRINNSTSELRLVPTAFEEGSAKLEKAEIOWTVHLTK
LOATGCKYEPANHALADPMSIGRECIETVARLTSVEGLDPSDPLRVRYSHLTK
ASOREANVMTSSMHHHPHPLPHHMAAFHPLPALQPNLHTESTEPCSLSTSEY
PSAGSALUTATFPAHADSALRMPSGGVAPCPVPLSTLSATVHAANAATPAAH
SFPLSFAGAPFMLPSNMAAANAATAAISPLSVSAASPOQSTGTNNKPYDPWGT
EVGAF"

```

BASE COUNT 715 a 604 c 553 g 667 t

ORIGIN

alignment_scores: Quality: 114.50 Length: 311
Ratio: 0.774 Gaps: 10
Percent Similarity: 47.588 Percent Identity: 21.543

alignment_block:

US-09-389-000-2 x MMU271867 ..

Align seg 1/1 to: MMU271867 from: 1 to: 2539

```

57 ProLeuThrMetGluLysMetSerAsnValValLysTyrTyrThrThr 73
    |||::: |||::: |||::: |||
100 CCGCTTGGGCGACAGTAGCTGCTCTCTGTCGCGATGAAGCGCCCTTG 149
    |||::: |||::: |||::: |||
73 sProSerAsnThrValLysThrGluAsnAlaThrGlyProGluGluLeu 90
    |||::: |||::: |||::: |||
150 TGAGGAACGACCTCCGAAAGCAGCTG.....GACGAGACCATGC 190
    |||::: |||::: |||::: |||
90 LysLeuProLeuGlnArgSerTyrSerGluHisLeuGlyTyrPheProThr 106
    |||::: |||::: |||::: |||
191 ACCTGGGAGGACGAGAACATTCCTGGGCGAC..... 222
    |||::: |||::: |||::: |||
107 AspLeuPheAlaCysSerGluSerLeuArgAsnGlyAsnGlyLeuGlu 123
    |||::: |||::: |||::: |||
223 .....GCTACAGCTCAGTGAAGCGCTCCCAATTCACCGACAC 260
    |||::: |||::: |||::: |||
123 uAsnAlaSerLeuSerGluPheGluLysAsnLysLysLysLeuLeu 140
    |||::: |||::: |||::: |||
261 TACCTCTCAGATTATGCAAGAAAGAAAGAGAGGATCATR..... 303
    |||::: |||::: |||::: |||
140 IsSerSerLysGluLysLeuArgArgGluArgGlyLeuTyrCysCysGlu 156
    |||::: |||::: |||::: |||
304 .....GAAAGGCGCTCGGATCGATGATATAACAATTATCT 342
    |||::: |||::: |||::: |||
157 GluLeuArgThrLeuLeuProTyrValLysGlyArgLysAsnAspAla.. 172
    |||::: |||::: |||::: |||
343 GAATTAGAGAGACTAGTGCACACCTTTTGAAGAAAGAGAGATCTCCAA 392
    |||::: |||::: |||::: |||
173 .....AlaSerValLeuGluAlaThrValAspTyrValLysTyrI 186
    |||::: |||::: |||::: |||
393 GTTAGAAAAGGCTGAATATGCAAAATGACAGTGAATCATTTGAAGATGC 442
    |||::: |||::: |||::: |||
186 LeArgGluLysIleSerProAlaValMet.....AlaGlnIleThr 199
    |||::: |||::: |||::: |||
443 TCAGGCTACAGGGGGAAGGCTACTTGTATGCCATGCTGTGCACA 492
    |||::: |||::: |||::: |||
200 GluAlaLeuGlnSerAsnMetArgPheCys.....LysLysGlu 212
    |||::: |||::: |||::: |||
493 GACCTCATGAGCATGTGATCCAGAGAGTGCCTGACAGAAAGTGCCTGTA 542
    |||::: |||::: |||::: |||
212 nglnthProIIeGluLeuSerLeuProGlyThrValMetAlaGlnArg 229
    |||::: |||::: |||::: |||
543 CTAAGCTCAGTGAAGGCGCTTGACCCGTGAGACCCACGCGTGGCC 592
    |||::: |||::: |||::: |||

```



```
|||||: : ||||| : : : : :
620 ...GAGGCGAGTGTGATCCATCTCCATGGCCACACCATCACCCCT 665
279 LysProTyrLys.....ThrHisIleSerSerAlaAla 290
      |||: : |||: : |||: : |||: :
666 TGCACCCCTCACACTGGGCGAGCTGCTTCCACCATCTCCCCACAGCCCTG 715
291 LeuSerLeuAsnSerLeuHisThrValArgTyrTyrSerLysValThrPr 307
      |||: : |||: : |||: : |||: : |||: :
716 CTCAGGCCCAATGAGCTCCACACA.....TCAGAGTCAACCCC 753
307 oSerTyrAspAlaThrAlaValThrAsnGlnAsnIleSerIleHisLeuP 324
      | : : : : :
754 ATGTCGCTATCCACATCT..... 772
324 roSerAlaMetProPheValSerSerPheSerLeuGlyThrAlaLeuLeu 340
      || : : : : : || : : : : : |||: : |||: :
773 .TCAGAGTGCCTTCTGCTCATGGCTGCTCTCTCCTCACAGCAGCTT 820
341 GlyTrpAlaArgArgAlaLeuHisIleProThr 351
      : : : ||| : : : : : : : : :
821 GCCCATGCAGATCTGCTCTTCGATGCCATCA 853
```

THIS PAGE BLANK (USPTO)

OM of: US-09-389-000-2 to: N_Geneseq_36: * out_format : pfs
Date: Jan 19, 2001 5:12 PM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frim+ p2n.model -DEV=xlp
-O=cn2_1/uspro.spool/US09389000/runat_17012001_152619_25058/app_query.fasta_1.467
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rmg -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosome2
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09389000 -CGEN1_1_133 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-389-000-2
Query length: 405
Database: N_Geneseq_36: *
Database sequences: 480022
Database length: 187831343
Search time (sec): 88.730000

score_list:

Sequence	Strid	Orig	ZScore	Escore	Len	Documentation
/SID56/gcgdata/geneseq/geneq/N1999.DAT:294275	+		2095.00	3660.08	4.4e-196	2128
/SID56/gcgdata/geneseq/geneq/NA2000.DAT:257039	+		115.00	186.33	0.0137	568
/SID56/gcgdata/geneseq/geneq/NA1999.DAT:227531	+		113.50	164.07	0.2379	3045
/SID56/gcgdata/geneseq/geneq/NA1999.DAT:Q75165	+		113.00	150.05	1.44	9370
/SID56/gcgdata/geneseq/geneq/NA1995.DAT:Q75164	+		113.00	150.02	1.44	9391
/SID56/gcgdata/geneseq/geneq/NA1996.DAT:T11413	+		109.50	163.80	0.2461	1704
/SID56/gcgdata/geneseq/geneq/NA1994.DAT:Q73000	+		108.50	161.03	0.3510	1857
/SID56/gcgdata/geneseq/geneq/NA1997.DAT:T70133	+		108.00	174.25	0.0567	510
/SID56/gcgdata/geneseq/geneq/NA1997.DAT:T70133	+		108.00	174.25	0.0612	537
/SID56/gcgdata/geneseq/geneq/NA1997.DAT:T70132	+		107.00	165.36	0.2016	1023
/SID56/gcgdata/geneseq/geneq/NA1997.DAT:T70130	+		105.00	176.56	0.0479	290
/SID56/gcgdata/geneseq/geneq/NA1993.DAT:Q38720	+		103.00	165.84	0.1895	537
/SID56/gcgdata/geneseq/geneq/NA1997.DAT:T43383	+		103.00	146.27	2.33	2866
/SID56/gcgdata/geneseq/geneq/NA1998.DAT:V27865	+		103.00	134.45	10.61	7881
/SID56/gcgdata/geneseq/geneq/NA1999.DAT:X18940	+		102.00	144.18	3.05	2948
/SID56/gcgdata/geneseq/geneq/NA1997.DAT:T63679	+		101.50	151.69	1.16	1437
/SID56/gcgdata/geneseq/geneq/NA1999.DAT:X36619	+		101.50	143.44	3.35	2911
/SID56/gcgdata/geneseq/geneq/NA1992.DAT:X26068	+		101.50	140.32	5.00	3804
/SID56/gcgdata/geneseq/geneq/NA1992.DAT:Q21097	+		101.00	150.11	1.42	1526
/SID56/gcgdata/geneseq/geneq/NA1998.DAT:V24018	+		101.00	147.90	1.89	1844
/SID56/gcgdata/geneseq/geneq/NA1997.DAT:T73287	+		100.00	139.64	5.45	3214
/SID56/gcgdata/geneseq/geneq/NA1997.DAT:T63679	+		100.00	139.64	5.45	3214
/SID56/gcgdata/geneseq/geneq/NA1998.DAT:V40019	+		99.50	141.79	4.14	2481
/SID56/gcgdata/geneseq/geneq/NA1998.DAT:V40019	+		99.50	141.79	4.14	2481
/SID56/gcgdata/geneseq/geneq/NA1999.DAT:Z10930	+		99.50	138.88	6.02	3183
/SID56/gcgdata/geneseq/geneq/NA1999.DAT:X26611	+		98.00	116.80	102.15	16788
/SID56/gcgdata/geneseq/geneq/NA1999.DAT:X13132	+		97.50	140.32	5.00	2080
/SID56/gcgdata/geneseq/geneq/NA1999.DAT:V72220	+		97.50	140.32	5.00	2100
/SID56/gcgdata/geneseq/geneq/NA2000.DAT:A34739	+		97.50	140.21	5.07	2100
/SID56/gcgdata/geneseq/geneq/NA2000.DAT:A34751	+		97.50	131.37	15.76	4475
/SID56/gcgdata/geneseq/geneq/NA2000.DAT:A34741	+		97.50	129.37	20.37	5312
/SID56/gcgdata/geneseq/geneq/NA1994.DAT:Q58522	+		97.50	129.37	20.62	5355
/SID56/gcgdata/geneseq/geneq/NA1999.DAT:X89220	+		97.50	129.37	20.62	5355
/SID56/gcgdata/geneseq/geneq/NA1999.DAT:X88221	+		97.50	129.37	20.62	5355
/SID56/gcgdata/geneseq/geneq/NA2000.DAT:Z39404	+		97.50	129.27	20.62	5355
/SID56/gcgdata/geneseq/geneq/NA2000.DAT:Z39405	+		97.50	129.27	20.62	5355
/SID56/gcgdata/geneseq/geneq/NA1995.DAT:Q75977	+		97.50	127.65	25.38	6151
/SID56/gcgdata/geneseq/geneq/NA1997.DAT:T46184	+		96.00	139.99	5.22	1707
/SID56/gcgdata/geneseq/geneq/NA1999.DAT:X91990	+		95.50	130.39	17.87	3600

/SID56/gcgdata/geneseq/geneq/NA1999.DAT:X20249 - 95.50 89.79 3.3e+03 116277
/SID56/gcgdata/geneseq/geneq/NA1996.DAT:T08974 + 95.00 128.53 22.69 3915
/SID56/gcgdata/geneseq/geneq/NA1999.DAT:X15635 + 95.00 128.53 22.69 3915
/SID56/gcgdata/geneseq/geneq/NA2000.DAT:A12989 - 94.50 142.89 3.60 1062
/SID56/gcgdata/geneseq/geneq/NA1999.DAT:X58669 + 94.50 137.98 6.75 1617

seq_name: /SID56/gcgdata/geneq/geneq/NA2000.DAT:294275

seq_documentation_block:

ID 294275 standard; cDNA; 2128 BP.

AC 294275;

DT 03-JUL-2000 (first entry)

DE Human PHEIX CDNA clone GPC12.

KW PHEIX; human; testis-specific; transcription factor;

KW prostate cancer; bladder cancer; ovary cancer; testicular cancer;

KW gene therapy; diagnosis; vaccine; chromosome 13q13.1-13.3; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 733..1953

FT /tag= a

PN WO200012709-A2.

PD 09-MAR-2000.

PF 31-AUG-1999; 99WO-US20137.

PR 31-AUG-1998; 98US-0098610.

PR 31-OCT-1998; 98US-0106524.

PT (UROC-) UROGENESIS INC.

PA (AFAR/) AFAR D E.

PA (HUBE/) HUBERT R S.

PA (RAIT/) RAITANO A B.

PI Afar DE, Hubert RS, Raitano AB;

WPI: 2000-237872/20.

P-PSDB: Y79269.

PT Testis specific Helix Loop Helix proteins expressed in cancers and

PT useful for the prevention, diagnosis and treatment of prostate, bladder

PT and ovarian tumors -

PT Claim 5; Fig 2A-D; 62pp; English.

CC This sequence is that of cDNA encoding human PHEIX (see Y79269),

CC a novel basic Helix Loop Helix protein thought to act as a

CC transcription factor. PHEIX normally exhibits a testis-specific

CC expression pattern but is up-regulated in prostate and other types

CC of cancer. The cDNA clone, termed GPC12 (ATCC 98956), was

CC isolated from a normal testis cDNA library using a suppression

CC subtractive hybridization method. The gene maps to chromosome

CC 13q13.1-13.3. The invention provides diagnostic and therapeutic

CC methods useful in the management of various cancers which express

CC PHEIX, including prostate cancer, bladder cancer, ovarian cancer

CC and testicular cancer, including therapies aimed at inhibition the

CC transcription, translation, processing or function of PHEIX. The

CC expression pattern of PHEIX suggests that is an ideal target for a

CC cancer vaccine approach to prostate cancer.

SQ Sequence 2128 BP; 675 A; 463 C; 464 G; 526 T; 0 other:

alignment_scores: Quality: 2095.00 Length: 405

Ratio: 5.173 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

Alignment block:
US-09-389-000-2 x 294275 ..

Align seg 1/1 to: 294275 from: 1 to: 2128

```

1 MetValLeuLeuLysValProSerSerLeuSerAlaGluGluLeuGluAl 17
|||||
733 ATGGTTCCTTGAAAGGCGCTTCTTCACTAACTGCGAGAGAGCTGGAGGC 782
17 AileuLeuLeuLeuArgPheGlyLysLysLysAsnThrHisSerLeuPheV 34
|||||
783 CATCAAGTAATTAATGATTGGCAAAAAGAAAATAACATTACCTGCTTG 832
34 AlPheIleIleProGluAsnPhelGlyCysIleSerGlyHisGlyMet 50
|||||
833 TTTTATATATCCCTGAAATTTTAAAGGTTGTATTTCAGGGCATGGAATG 882
51 AspIleAlaLeuThrGluProLeuThrMetGluLysMetSerAsnValVa 67
|||||
883 GATATGCTTTAACTGAACCACTGACAAATGAAAAAATGATTAATGTGT 932
67 LlystyTrpThrCysProSerAsnThrValIlysthrGluAsnAlat 84
|||||
933 AAAATACAGCAACATGTCCCTCAACACAGTGTAAAGCTGAAAAAGCAA 982
84 hrgLysProGluGluLeuGlyLeuProLeuGlnArgSerTyrSerLysHis 100
|||||
983 CTGGGCGTGAAGAACTTGATGGCTCCCTGACAGAGTCTTACAGCAGCAAC 1032
101 LeuGlyTyrPheProThrAspLeuPheAlaCysSerGluSerLeuArgAs 117
|||||
1033 CTGGATATTTTCTTCTACTGATCTATTGCTGCTGCTGATCTTTAAGGA 1082
117 nGlyAsnGlyLeuGluLeuAsnAlaSerLeuSerGluPheGluLysAsnL 134
|||||
1083 TGGCAATGGGCTTGAATTAATGCTTCCTTGCAAGAGTTCCGAAAAACA 1132
134 LysLysIleSerLeuLeuHisSerSerLysGluLysLeuArgArgGluArg 150
|||||
1133 AAAAGATCTCTCTTCTTCAATTCACAGAAAGAAAACTAAGAAAGGAAAA 1182
151 IleLysTyrCysCysGluGluLeuArgThrLeuLeuProTyrValIlysgl 167
|||||
1183 ATCAAAATTTCTGTGAGCAGCTGCTGCTGCTGCGGTATGTAAGAAAG 1232
167 YArgLysAsnAspAlaAlaSerValLeuGluAlaThrValAspTyrValL 184
|||||
1233 GAGAAAGAAATGATGGGCTTCACTTCTTGAGGCAACAGTTGATATGTGA 1282
184 YstYrIleArgGluLysIleSerProAlaValMetAlaGluIleThrLys 200
|||||
1283 AATATATCCGGAGAAAATCTCTCCAGCGGTATGCGCCAGATTACAGAA 1332
201 AlaLeuGlnSerAsnMetArgPheCysLysLysGlnGlnThrProIleGl 217
|||||
1333 GCACCTTCAGACCAACATGAGGTTTGTGAAGAAACACAAACACCATGGA 1382
217 uLeuSerLeuProGlyThrValMetAlaGlnArgGluAsnSerValMetS 234
|||||
1383 GCTGTCTCTCCAGGCACTGTCACTGCGACACGGGAAAAACAGTGCATGA 1432
234 eTrhYrSerProGluArgGlyLeuGlnPheLeuThrAsnThrCysTrp 250
|||||
1433 GCACCTTACTCCCTGAGAGAGGCTCCCAATTCCTGACTAAATACGCTGG 1482
251 AsnGlyCysSerThrProAspAlaGluSerSerLeuAspGluAlaValAr 267
|||||
1483 AATGGTGCTCTCACTCTGATGACAGAGAGCTCTTGATGAAGCTGTGAG 1532
*267 gValProSerSerSerAlaSerGluAsnAlaIleGlyAspProTyrLysT 284
|||||

```

```

1533 AGTTCATCAAGCTCCGCCCTCAGAGAAATGCTATTGATGATCATATATAA 1582
284 hrHisIleSerSerAlaAlaLeuSerLeuAsnSerLeuHisThrValArg 300
|||||
1583 CTCACATTTCCAGTGACAGCGCTGCTGGAATTCCTTGCAATCATGTGAGA 1632
301 TyrTrpSerLysValThrProSerTyrAspAlaThrAlaValThrAsnGl 317
|||||
1633 TATTATTCTAAAGTCACCCCTTCTTACGATGATCACTGCTGAACAAATCA 1682
317 nAsnIleSerIleHisLeuProSerAlaMetProValSerSerPheS 334
|||||
1683 GAACATTTCAATTCATTACCTTCAGCCATGCCCGGCTCAAGCTTCT 1732
334 eTrLeuGlyThrAlaLeuLeuGlyTrrPalaArgAlaLeuHisIlePro 350
|||||
1733 CCCTCGGCAGTCACCTTGGGTTGGGCCAGACGCTGCACTACACATCCCA 1782
351 ThrValCysAsnSerPheGlyArgIleLysSerThrCysLeuLysPheTh 367
|||||
1783 ACTGTCTGCAACAGTTTGGGCGTATTAAAGCAACATGTTGAATTCAC 1832
367 rLeuSerThrThrTyrTrpAlaGlnPheAspAsnLeuGlyLysValGlu 384
|||||
1833 ACTCTCAACCACTACTGGCGCAGTTTGACATCTAGAAAAGTGGAAC 1882
384 lnaArgMetIleLeuLysAlaProProLysAspLeuIleSerLysGluLeu 400
|||||
1883 AAAAGATGATTTTGAAGCTCCACCAAAAGCACTTAATATCAAAAGAGTTG 1932
401 AlaTrpPheGlyPhe 405
|||||
1933 GCATGGTTGGCTTC 1947

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT.Z57039
seq_documentation_block:
ID Z57039 standard; CDNA: 568 BP.
XX
AC Z57039;
DT 19-MAY-2000 (first entry)
XX
DE PBR-associated protein (PAP)8 encoding cDNA.
XX
KW Peripheral-type benzodiazepine receptor; PBR. PBR-associated protein;
KW PAP; cell proliferation; cancer; cell death; cytosolic; neuroprotective;
KW immunomodulator; antiinfectility; cerebroprotective; atherosclerosis;
KW Niemann-Pick C; tumour; Alzheimer's disease; developmental disorder;
KW cholesterol; multiple sclerosis; stress; neurodegenerative disorder;
KW immune disorder; stroke; PAP8; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 1..477
FT FT /*tag= a
FT FT /product= "PAP8"
XX
XX WO200009549-A2.
XX
PD 24-FEB-2000.
XX
PF 11-AUG-1999: 99WO-US18507.
XX
PR 11-AUG-1998: 98US-0096048.
XX
XX (GEOU) UNIV GEORGETOWN MEDICAL CENT.
XX
XX Papadopoulos V, Li H;
XX
XX WPI: 2000-224278/19.
XX
XX P-PSDB: Y67501.
DR

```



```

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1996.DAT:T11413
357 yargilelysserthr 362
|||||: |||||:
4012 CAGGGTTTCAGCAGC 4027

seq_documentation_block:
ID T11413 standard; DNA; 1704 BP.
XX AC T11413;
XX TT
XX 02-JUL-1996 (first entry)
XX
XX Sequence encoding maltose converting enzyme.
XX KW Maltose; trehalose; sugar; sweetener; food industry; foodstuffs;
XX ss.
XX OS Plasmobacter sp. R48.
XX
XX Key Location/Qualifiers
FH FT 1..1704
CD S /*lag= a
/*product= Maltose converting enzyme.
XX
XX EP695804-A2.
XX
XX 07-FEB-1996.
XX PD
XX PF 08-JUN-1995; 95EP-0303933.
XX PR 16-JUN-1994; 94JP-0156399.
XX PA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.
XX PI Kubota M, Sugimoto T, Tsusaki K;
XX DR WPI: 1996-088926/10.
XX P-PSDB: R89869.
XX
XX Recombinant enzyme which converts maltose to trehalose and vice
PT versa - has mol. wt. of 57.67 kilodaltons and isoelectric point of
PT 4.1-5.1, useful for industrial scale prodn. of trehalose
XX PS
XX Claim 4; Page 21; 31pp; English.
XX
XX The recombinant enzyme encoded by this sequence converts maltose to
CC trehalose. The trehalose produced can be used as a sweetener in the
CC food industry because it contains no reducing residues and so does
CC not cause a browning reaction even when heated in the presence of
CC amino acids. Conversion of maltose to trehalose preferably occurs
CC when an effective amount of the enzyme is allowed to co-exist in an
CC aqueous medium containing maltose at up to 50% w/v to catalyse the
CC reaction at a temp. of 4-45 degrees Celsius and a pH of 5.5-9.0. A
CC composition comprising trehalose up to about 80% w/w is obtained.
XX CC
XX Sequence 1704 BP; 274 A; 611 C; 556 G; 263 T; 0 other:

Alignment_scores:
Quality: 109.50 Length: 351
Ratio: 0.730 Gaps: 14
Percent Similarity: 42.735 Percent Identity: 20.513

alignment_block:
US-09-389-000-2 x T11413 ..
Align seg 1/1 to: T11413 from: 1 to: 1704

44 CysIIesrgYIHsGLyMeCaSPILeAlaLeuThrcIuPro..... 57
||||| ||||| ||| ::: :::: ||| |||
680 TGCTCAAGGCGGTGCGGCTTGCTGCACGACACTAACCCGACCGGGTg 729
```

[illegible]

||||| |||||
 1384 CACATCCGCCGACGCGACGCTTCGGG..... 1413

364 ulyspetherleuserthrtrtyrtrpalaaglnpheaspaslnleugly 380
 1414CTGGGCACC.....TTCGAGGACCTCGGC 1437

seq_name: /SIDS6/gcdata/geneseq/geneseqn/NA1994.DAT:Q73000

seq_documentation_block:
 ID Q73000 standard; cDNA; 1857 BP.

AC Q73000;

DT 08-JUN-1995 (first entry)

DE Petunia ph6 gene.

KM Vacuolar pH; pH gene; Petunia; ss.

OS Petunia hybrida strain V26.

XX Key Location/Qualifiers

FT CDS 481..1632

FT misc_feature 484..1629

FT /tag= b

FT /label= claimed

PN MO9423561-A.

XX 27-OCT-1994.

XX 15-APR-1994; 94WO-US04173.

XX 16-APR-1993; 93US-0049282.

PA (DNAP) DNA PLANT TECHNOLOGY CORP.

XX Chuck GS, Courtney-guterson N, Dooner HK, Keller J;

PI Mijar CS, Ralston EJ;

XX MPI: 1994-341349/42.

DR P-PSDB: R62656.

XX Petunia Ph gene and constructs containing it - for alteration of

PT vacuolar pH used in the formation of blue flowers

XX Claim 7; Page 42-44; 62pp; English.

XX The V26 strain of Petunia was used. Poly A RNA was isolated from
 CC total RNA from flower buds and used to generate a cDNA library in
 CC the vector lambda Zapit (Stratagene). The SstI to BamHI fragment at
 CC the left hand side of Ac was used to isolate ppet14-1 (cont. p6
 CC cDNA), the sequence for which is claimed (see Q73000 FT). The AA
 CC sequence comprises a helix-loop-helix structural motif starting with
 CC the sequence NHYLAR (starting at residue 192) and extending to the
 CC sequence KKVDLE (ending at residue 245). Proteins including this
 CC motif include the myc family of oncogenes, regulators of neuron and
 CC muscle development, and regulators of segmentation and organ
 CC patterning in Drosophila. In plants, the motif is found in R(S),
 CC a protein involved with regulating anthocyanin synthesis in maize.
 CC Comparison of the p6 and R(S) sequences over the 54 AA helix-loop-
 CC helix region detected 57% AA identity. Outside this region only 19%
 CC identity was found.

XX Sequence 1857 BP; 565 A; 415 C; 413 G; 464 T; 0 other;

alignment_scores:

Quality: 108.50 Length: 306

Ratio: 0.714 Gaps: 15

Percent Similarity: 49.673 Percent identity: 23.529

alignment_block:

US-09-389-000-2 x Q73000 ..

Align seg 1/1 to: Q73000 from: 1 to: 1857

6 ValProserSerLeuSerAlaGlnGluLeuGluAlaIleLysLeuIleAr 22
 :|||:|||||
 670 TTACCAAGGCGGCGCTAGTATGATGATTA..... 699
 22 gphesglyLysLysAsnThrHis...SerLeuPheValPheIleIleP 38
 :|||:|||||
 700TCACAAGAAGACACACATTACTCTCAACAGAGTGCACCATTC 742
 38 roGluAsnPhelLys.....Gly 43
 :|||:|||||
 743 TTGAACACCTCTCAACCAACCTCAAAATTTTCTCTACCAATAATGGCG 792
 44 CysIleSerGlyHisGlyMetAspIleAlaLeuThrGluProLeuThrMe 60
 :|||:|||||
 793 TGTATTTCC..... 801
 60 tGluLysMetSerAsnValIleLysLysTrpThrCysProSerAsnT 77
 :|||:|||||
 802 .CAACAACCCAACTGCGCTTCAACAAGTGGCCAGC...CCACACACA 847
 77 hValLysThrGluAsnAlaThrGlyProGluGluLeuGlyLeuProLeu 93
 :|||:|||||
 848 CCGTCTCCAGCCCATTTCTTGACGGCGGCGCCACCTCGGCGCAGTGCGTG 897
 94 GlnArgSerTyrSerGluHisLeuGlyLysTrpPheProThrAspLeuPheAl 110
 :|||:|||||
 898 CTCAAAGCATCTATTCTCTGTCTTCCATTCTTCACACTAATACCACAC 947
 110 acYsSerGlu...SerLeuArgAsnGlyAsnGlyLeuGluLeuAsnAla 126
 :|||:|||||
 948 TGCAGCTGAAGTTCTCCAAAGTCACAGTCAGCGCTACACTGTGATTCCT 997
 126 er...LeuSerGluPheGluLysAsnLysLysIle..... 136
 :|||:|||||
 998 CCAGTCATCTCGCTTCGAAAGGCTAGTATTAACAAGAACAGAGCCT 1047
 137 SerLeuLeuHisSerSerLysGluLysLeuArgArgGluArgIleLysTy 153
 :|||:|||||
 1048 AGTGAACACATGATCTGCTGACAGCAGCGCTAGAGAAAGCTCAACGA 1097
 153 rCysCysGluGlnLeuArgThrLeuLeuProTyrValLysGluArgLysA 170
 :|||:|||||
 1098 ACGGTTTATCATTTGAGGTGCTGTCCTTTGTT.....ACGAAA 1141
 170 snAspAlaIleSerValLeuGluAlaThrValAspTyrValLysTyrIle 186
 :|||:|||||
 1142 TGGATTAAGCCCTCCATTTGTTGTCACACCATTAAGATTAAGTGA 1191
 187 ArgGluLysIleSerPro...AlaValMetAlaGlnIleThrGluAla 202
 :|||:|||||
 1192 CGTAAAGAAAGTTCAAGATCTTGAAGCTAGAGCCATCAAGCAGAGCT. 1239
 202 uGlnSerAsnMetArgPheCysLysGlnGlnIleThrProIleGluLeu 219
 :|||:|||||
 1240ACGCTGCAGACAA 1252
 219 erLeuProGlyThrValMetAlaGlnArgGluAsnSerValMetSerThr 235
 :|||:|||||
 1253 AGGATACAGGACTGTGAAGGTGTGCAAGAGAGGGT..... 1290
 236 TyrSerProGluArgGlyLeuGlnPheLeuThrAsnThrCysTrpAsnG 252
 :|||:|||||
 1291AAGAGGAGAATGACATAGTCGAAGAGAAATGTTGGTGAGG 1331
 252 Y.....CysSerThrProAspAlaGlnSerSerLeuAspG 264
 :|||:|||||
 1332 ACAGCAAAAGATCAAGCATCTTCGCCCTCAACGACACATGAA...GAGG 1378


```

PR 01-JUN-1994: 94US-0252966.
PR 19-SEP-1991: 91US-0756195.
PR 23-JUN-1992: 92US-0903710.
PR 01-APR-1994: 94US-0222638.
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PI Ayer DE, Eisenman RN:
DR WPI: 1997-258216/23.
DR P-PSDB: W10041.
XX msin nucleic acids encoding recombinant polypeptide(s) that
PT associate with Mad polypeptide - are possible homologues of S.
XX cerevisiae general repressor protein
XX
XX Example 2; Fig 2A; 11bp; English.
XX
CC This CDNA sequence, encodes Max 11, a basic helix-loop-helix zipper
CC (bHLHZip) protein. Max 11 and Max 14 (T70134) appear to be partial,
CC overlapping CDNs. Subsequent isolation of several overlapping CDNs
CC from a Mencia (human Burkitt's lymphoma cell line) lambda gt10 library
CC permitted deduction of an apparently complete open reading frame for
CC Max that encodes 151 residues. The 9-amino acid insertion found in
CC several PCR clones is not indicated in the specification. Max is an
CC obligate partner for the DNA binding and transcriptional functions of
CC Myc family proteins as well as for the Mad protein. Max is a stable,
CC ubiquitously expressed protein which in general does not appear to be
CC regulated during mitogenesis, the cell cycle, or differentiation.
CC Expression of Mad is closely linked to differentiation in at least two
CC distinct cell lineages. The switch from Myc:Max to Mad:Max complexes may
CC reflect the repression of transcription of Myc regulated genes by Mad.
CC The DNA, vectors and host cells of the invention are useful for the
CC recombinant production of msin proteins useful in elucidation of Mad
CC repressor functions.
XX
SQ Sequence 537 BP; 164 A; 154 C; 141 G; 78 T; 0 other;

```

```

Alignment_scores:
  Quality: 108.00      Length: 131
  Ratio: 1.301        Gaps: 3
  Percent Similarity: 63.359   Percent Identity: 24.427

```

```

Alignment block:
US-09-389-000-2 x T70133 ..

```

```

Align seg 1/1 to: T70133 from: 1 to: 537

```

```

140 HsSerSerLysGluLysLeuArgArgLysGluArgLysLysLysCysGsgl 156
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
109 CATATGCACTGAGAGAAAGCTAGGAGCCATCAAGACAGCTTCA 158
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
156 uGlnLeuArgThrLeuLeuProTyrValLysGlyArgLysAsnAspAlaA 173
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
159 CAGTTGGCGGACTCAGTCCCATCTCCAGAGAGAGAGGATCCCGG 208
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
173 laserValLeuGluAlaThrValAspTyrValLysTyrLleArgLys 189
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
209 CCCAAATCTGACAAAGCCACAGAGTATATCCATATATGGAAGAA 258
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
190 lIeSerProAlaValMetAlaGlnIleThrGluAlaLeuGlnSerAsn 206
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
259 .....AACCAACACACACACCAAGATAT 281
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206 tArgPheCysLysLysGlnGlnThrProLleGluLeuSerLeuProGlyT 223
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
282 TGACGACCTCAAGCGGAGATGCTCTTGAGAGCAAGACCTGTCAC 331
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
223 hrValMetAlaGlnArgLysAsnSerValMetSerThrTyrSerProGlu 239
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
332 TGGAGAAAGCG...AGGTCAAGTGGCCCACTGACAGCAACCACTACCCCTCC 378

```

```

240 ArgGlyLeuGlnPheLeuThrasn.....ThrcYstr 250
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
379 TCAGACACACAGCCTCTACACCAAGCCCAAGGACAGACACATCTGCGCTT 428
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
250 pAsnGlyCysSerThrProAspAlaGluSerSerLeuAspGlu 264
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
429 CGATGGGGGCTCAGACTCCAGCTCAGACTGAGCTGAGCTGAGAG 471
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: /SIDS6/gcdata/geneseq/geneseqn/NA1997.DAT:T70132
seq_documentation_block:
ID T70132 standard; CDNA to mRNA: 1023 BP.
XX
XX T70132;
AC
XX 07-FEB-1998 (first entry)
DT
XX
XX Max-Interacting protein coding sequence (clone 20).
DE
XX
XX murine; mSinA; mammalian homologue; Saccharomyces cerevisiae; repressor;
KW Sin3; Mad; Max; mSin:Mad complex; mSin:Mad:Max complex; Myc; promoter;
KW basic helix-loop-helix zipper protein; compete; DNA-binding;
KW Myc:Max complex; activate; transcription; gene regulation; ss.
XX
XX Mus musculus.
XX
XX US5624818-A.
XX
XX 29-APR-1997.
PD
XX
XX 01-JUN-1994: 94US-0252966.
PF
XX
XX 01-JUN-1994: 94US-0252966.
PR 19-SEP-1991: 91US-0756195.
PR 23-JUN-1992: 92US-0903710.
PR 01-APR-1994: 94US-0222638.
XX
XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PA
XX
XX Ayer DE, Eisenman RN:
PI
XX
XX WPI: 1997-258216/23.
DR
XX
XX msin nucleic acids encoding recombinant polypeptide(s) that
PT associate with Mad polypeptide - are possible homologues of S.
XX cerevisiae general repressor protein Sin3
XX
XX Example 13; Fig 27; 11bp; English.
XX
XX This CDNA sequence, designated clone 20, encodes a novel murine basic
XX helix-loop-helix zipper (bHLHZip) protein which interacts with the
XX bHLHZip, Max. Max is an obligate partner for the DNA binding and
XX transcriptional functions of Myc family proteins as well as for the Mad
XX protein. Max is a stable, ubiquitously expressed protein which in general
XX does not appear to be regulated during mitogenesis, the cell cycle, or
XX differentiation. Other CDNA sequences encoding murine proteins (msin)
XX which may be mammalian homologues of the Saccharomyces cerevisiae general
XX repressor protein Sin3 are claimed. The mSin proteins associate with Mad
XX and Max to form a mSin:Mad:Max complex which binds to a CAGGTG sequence
XX in promoters. Mad:Max complexes repress, while Myc:Max complexes
XX activate, transcription from promoters containing proximal CAGGTG binding
XX sites for these proteins. Expression of Mad is closely linked to
XX differentiation in at least two distinct cell lineages. The switch from
XX Myc:Max to Mad:Max complexes may reflect the repression of transcription
XX of Myc regulated genes by Mad. The DNA, vectors and host cells of the
XX invention are useful for the recombinant production of mSin proteins
XX useful in elucidation of Mad repressor functions.
XX
SQ Sequence 1023 BP; 275 A; 293 C; 279 G; 176 T; 0 other;

```

```

Alignment_scores:
  Quality: 107.00      Length: 266

```



```

100 CGGGCCCGCCCAAGGCCCTACGGTGGCCGAGGTTCCAGC..... 140
136 lIeSerLeuLeuHisSerSer.....LysGluLysLeuArgGluArg 150
141 ...TCTCGCTTGGCGGAGAACGTGATCAGCAGAGAGGAGCGAGAGAGCG 187
150 gIleLysTyrCysGluGlnLeuArgThrLeuLeuProTyrValLysG 167
188 GATGTGCTTGAGCTGTGAGCTGTGCGGCGCTGCTGCCAGTTGATG 237
167 lYArgLysAsnAspAlaAlaSerValLeuGluAlaThrVal 180
238 GCCGGCGGAGAGACATGCGCTGCTGCTGAGATGCTGTT 278

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1993.DAT:Q38720

seq_documentation_block:
ID Q38720 standard; DNA; 537 BP.
XX
AC Q38720;
XX
DT 15-JUL-1993 (first entry)
XX
DE Max CDNA.
XX
KW Med; max. myc; c-myc; helix-loop-helix zipper; leucine zipper;
KW helix-turn-helix; diagnosis; prognosis; cancer; malignancy;
KW neoplasm; tumour; studying embryogenesis; study gene regulation.
XX
FH Key Location/Qualifiers
FT CDS 28..508
FT /tag= a
XX
XX W09305056-A.
XX
XX 18-MAR-1993.
XX
XX 09-SEP-1992; 92WC-US07629.
XX
XX 09-SEP-1991; 91US-0756195.
XX
XX 23-JUN-1992; 92US-0903710.
XX
XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
XX Ayer DE, Blackwood EM, Eisenman R;
XX
XX WPI; 1993-100913/12.
XX
XX P-PSDB; R33386.
XX
XX Helix-loop-helix zipper protein named Max - associates with MYC
XX PT or Mad polypeptide(s) useful as diagnostic or prognostic tools
XX PT for diverse types of cancer
XX
XX Claim 15; Fig 2; 93p; English.
XX
XX This sequence represents CDNA for Max protein. A random primed
XX lambda gt11 expression library was derived from a baboon lymphoid
XX cell line 5945 as described in [RL Idzerda et al. Proc. Natl. Acad.
XX Sci. USA. 86, 4659 (1989)]. Phage from this library produce nearly
XX full length beta-galactosidase proteins fused with the ORF of the
XX directionally cloned cDNAs. More than 10power4 plaques were
XX screened for their ability to interact with 125-I labeled
XX GST-Myc92. Several positive plaques were identified, and Max 11
XX and 14 survived multiple rounds of plaque purification. Sequence
XX analysis of these two clones together with those derived from a
XX Max lambda gt10 library enabled the sequence shown below to be
XX deduced.
XX
XX Sequence 537 BP; 165 A; 158 C; 133 G; 81 T; 0 other;
XX
alignment_scores:

```

```

Quality: 103.00 Length: 136
Ratio: 1.272 Gaps: 3
Percent Similarity: 59.559 Percent Identity: 25.735

alignment_block:
US-09-389-000-2 x Q38720 ..

Align seg 1/1 to: Q38720 from: 1 to: 537

140 HisSerSerLysGluLysLeuArgGluArgGluGluLysTyrCysGlu 156
109 CATATGCACTGAGAACGAAAGAGTGGAGGACCATCAAAAGACAGCTTCA 158
156 uGlnLeuArgThrLeuLeuProTyrValLysGluArgLysAsnAspAla 173
159 CAGTTGGCGGACTGATCCATCTCCTCAAGAGAGAGGAGCATCCAGG 208
173 lAsSerValLeuGluAlaThrValAspTyrValLysTyrIleArgGluLys 189
209 CCCAATCTTAGACAAAGCCACAGAGTATATCTACTATATGCGAAGANA 258
190 lIeSerProAlaValMetAlaGlnIleThrGluAlaLeuGlnSerAsnMe 206
259 .....AACCAACACACACAGCAAGATAT 281
206 tArgPheCysLysGluGlnThrProIleGluLeuSerLeuProGlyT 223
282 TGACAGCTCTCAAGCGGAGATGCTTCTGTGAGCAGCACTCCGTGAC 331
223 hrValMetAlaGlnArgGluAsnSerValMetSerThrTyrSerProGlu 239
332 TGGAGAAAGCG...AGTTCAGTGGCCCAATCCAGACCACTACCCCTCC 378
240 ArgGluLeuGlnPheLeuThrAsnThrCysTrpAsnGlyCysSerThrPr 256
379 TCAGACAACAGCTCTACCAAC..... 402
256 oAspAlaGluSerSerLeuAspGluAlaValArgValProSerSerSer 273
403 ...GCCAAGGCGACACCATCTCTGCTTTGATGGGGCTCAGACTCCA 448
273 lAsSerGlu 275
449 GCTCAGAG 456

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1997.DAT:T43383

seq_documentation_block:
ID T43383 standard; CDNA; 2866 BP.
XX
XX T43383;
XX
XX 11-MAR-1997 (first entry)
XX
XX Human cytokine response gene CR8.
XX
XX DE Human cytokine response gene CR8.
XX
XX KW Cytokine response gene; CR8; interleukin-2; IL-2;
XX KW ligand-stimulated gene expression; diagnosis; therapy;
XX KW transcription factor; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT CDS 240..1478
XX FT /tag= a
XX
XX W09639427-A1.
XX
XX PD 12-DEC-1996.
XX
XX 05-JUN-1996; 96WO-US09194.
XX
XX 05-JUN-1995; 95US-0465585.
XX

```


CC human colon tumour cell line LIM1863 cDNA library using mouse TRIP1
CC partial cDNA clone cml-85-g3 as probe to isolate overlapping clones,
CC and use of PCR to obtain additional 5' and 3' sequences. The
CC invention relates to novel genes encoding components of the
CC telomerase enzyme complex, including TRIP1 and telomerase protein 2
CC (TP2, see also W61349). Nucleic acids that code for such
CC polypeptides may be useful as therapeutic agents in those cases
CC where increasing TRIP1 activity or TP2 activity is desired, e.g.
CC for treatment of HIV infection, AIDS and ageing disorders. In
CC situations in which TRIP1 and/or TP2 activity is to be decreased,
CC such as in cancer cells in which TRIP1 activity and/or TP2 activity
CC is elevated, TRIP1 and/or TP2 may serve as a target to identify a
CC molecule which inhibits activity, or which decreases or inhibits
CC the protein-protein interaction of TRIP1 and TP2, or the binding of
CC either TRIP1 or TP2 to telomerase RNA. Alternatively, ex vivo or
CC in vivo gene therapy may be used to administer TRIP1 or TP2
CC antisense molecules, or DNA constructs may serve to disrupt or
CC enhance TRIP1 and/or TP2 expression in cells, and to create
CC dominant negative inhibitors of TRIP1 or TP2.

CC Sequence 7881 BP; 1716 A; 2199 C; 2221 G; 1745 T; 0 other;

alignment_scores:

Quality: 103.00 Length: 403
Ratio: 0.613 Gaps: 20
Percent Similarity: 41.687 Percent Identity: 21.092

alignment_block:

US-09-389-000-2 x V27865 ..

Align seg 1/1 to: V27865 from: 1 to: 7881

```

48 HisGlyMetAspIle.....AlaLeuThr..... 55
||||| ||||| ||||| |||||
4219 CACGGGCGCTGAGTCTTCCCGACGCTTGACCTGACCTGACGACACAG 4268
56 GluProLeuThrMetGluLysMetSerAsnValValLysTyrTrpThr 72
||| ||||| ||||| ||||| |||||
4289 GAGTGTTGACTGTGACACAGCTGACGAGAGTGTGTGCGGA 4318
72 hrCysProSerAsnThrValLysThrGluAsnAlaThrGlyProGlu 88
||| ||||| ||||| ||||| |||||
4319 CACATCCGAGGAGGACTTAAGACTGGAAGAAGAGTGTCTGTGTAAC 4368
89 LeuGlyLeuProLeuGlnArgSerTyrSerGluHisLeuGlyTyrPhePr 105
||| ||||| ||||| ||||| |||||
4369 AGTGAAGACCC.....TACCC 4385
105 cThrAspLeuPheAlaCys...SerGluSerLeuArgAsnGlyAsnGlyL 121
||| ||||| ||||| ||||| |||||
4386 CATGGCCCGCTTGGCTGCTGCTGCACAGTGTGCGC..... 4422
121 euGluLeuAsnAlaSerLeuSerGluPheGluLysAsnLysLysIleSer 137
||| ||||| ||||| ||||| |||||
4423 .....AGT 4425
138 LeuLeuHisSerSerLysGluLysLeuArgArgGluArgGlyLysTyrCy 154
||||| ||||| ||||| ||||| |||||
4426 TTGCTA.....GGGAGGCGCCCTTGAGAGCGCCCTGTCGCCGCTGTG 4469
154 scyGluGluIn.....LeuArgThrLeuLeuProTyrValLysGlyArgL 169
||| ||||| ||||| ||||| |||||
4470 CCTCCCTATGATGGCCCTGAGACACAGCAGCTTAACGTTGCTATGGGAAGA 4519
169 ys.....AsnAspAlaAlaSerValLeuGluAlaThrValAspTyr 182
||| ||||| ||||| ||||| |||||
4520 GGCACAGGCTAGAGACAGCAGCACATCTC..... 4551
183 ValLysTyrIleArgGluLysIleSerProAlaValMetAlaGlnIleTh 199
||| ||||| ||||| ||||| |||||
*4552 .....ATTGACGCTCAGCTC... 4566

```

```

199 rGluAlaLeuGlnSerAsnMetArgPheCysLysLysGlnGlnThrProI 216
4567 .....TGAGACATGTGACCTATGCTTACGCA 4597
216 leGluLeuSerLeuProGlyThrValMetAlaGlnArgGluAsnSerVal 232
||| ||||| ||||| ||||| |||||
4598 CCTTCGAAAGTTGGCCCTCTGAGGCTGTGGAGACCTGCTTACACACTG 4647
233 MetSerThrTyrSerProGluArgGlyLeu.....GlnPheLeuThrAs 247
||| ||||| ||||| ||||| |||||
4648 CTCCAG.....AGCGGAGCCGTGAGCTCTTTCGAAGTCTTACCA 4691
247 nThr..... 248
4692 COTCATGTGGTGGCTGCACACTTGAATTTGGTCTGTCTGCGCTCT 4741
249 .....CysTrpAsnGlyCysSerThrProAspAlaGluSerSer 261
||| ||||| ||||| ||||| |||||
4742 TGGAGCCCAAGCCCTATGCTTCTTCATGCTCCCAAGAGAACAAAG 4791
262 LeuAspGluAla..... 265
||| ||||| ||||| ||||| |||||
4792 CTCCCGAGGCTGACGTTGACGTGTTGCACTTCTGAGGACAGCAGCC 4841
266 .....ValArgValProSerSerSerAlaSerGluA 276
||| ||||| ||||| ||||| |||||
4842 TTCAATCTCAGCCAGTACCCCGGCTCTGCCCCAGAGCAGCAGCACC 4891
276 snAlaIleGlyAspProTyrLysThrHisIleSerSerAlaAlaLeuSer 292
||| ||||| ||||| ||||| |||||
4892 AGCCCTGAGCTACACCTCTTGCACACAGCCTGCGTCTCCCGGAGA 4941
293 LeuAsnSerLeuHisThrValArgTyrTyrSerLysValThrProSerTy 309
||| ||||| ||||| ||||| |||||
4942 TGGCACCTCAACACACACTGAGTGGCTTATATAA..... 4977
309 rAspAlaThrAlaValThrAsnGlnAsnIleSerIleHisLeuPro...S 325
||| ||||| ||||| ||||| |||||
4978 CCCCCGACCATGAATAATCAGCAAGCTCCAGCTGTCTGTGCGAGTTT 5026
325 erAlaMetProProValSerSerPheSer..... 334
||| ||||| ||||| ||||| |||||
5027 CCTATCCCTACGCTAGCTGTGCTTCTCCACCAATGGGCAAGAGCAGCT 5076
335 LeuGlyThrAla.....LeuLeuGlyThrPal 343
||| ||||| ||||| ||||| |||||
5077 GTGGGCACTGCCAATGGGACAGTTTACCTGTTGGACCTGAGACACTTG 5124
343 aArgArgAlaLeuHisIleProThrValCysAsnSerPheGlyArgIleL 360
||| ||||| ||||| ||||| |||||
5125 CAGAGAGAGAAGTCTGTGTGAGTGTGCTGTGATGGAATC..... 5163
360 ysSerThrCysLeuLysPheThrLeuSerThrThrTyrTrpAlaGlnPhe 376
||| ||||| ||||| ||||| |||||
5164 ..TCTGCTGTGTGCTCTCCGATATACACTTCTTACTGCTTC 5211
377 AspAsnLeu 379
||| ||||| ||||| ||||| |||||
5212 GACGGGCTC 5220

```

seq_name: /sids6/gcgdata/geneseq/geneseqn/NA199.DAT.X18940

seq_documentation_block:

ID_X18940 standard; CDNA: 2948 BP.

XX X18940;

AC X18940;

XX X18940;

DT 11-MAY-1999 (first entry)

XX XX

DE Human basic helix-loop-helix transcription-control factor gene.

XX Human; chondrocyte; basic helix-loop-helix transcription-control factor;

KW Chondrite differentiation; gene therapy; deformative arthritis;


```

KW detection; ds.
XX Homo sapiens.
XX Key Location/Qualifiers
FH CDS 207..1445
FT CDS /*tag= a
FT
XX
XX MO9902677-A1.
XX
XX 21-JAN-1999.
XX
XX 10-JUL-1998; 98WO-JP03106.
XX
XX 11-JUL-1997; 97JP-0202227.
XX
XX (CHUS ) CHUGAI SEIYAKU KK.
XX
XX Kato Y, Kawamoto T;
XX
XX MPI; 1999-120878/10.
XX
XX P-PSDB; W99060.
XX
XX New human chondrocyte-originated gene encoding basic
XX helix-loop-helix transcription-control factor - used for detection
XX of differentiated, studying chondrite differentiation and
XX degeneration mechanism to develop gene therapy for, e.g. deformative
XX arthritis
XX
XX Claim 2; Page 26-30; 37pp; Japanese.
XX
XX The present sequence encodes a basic helix-loop-helix transcription-
XX control factor isolated from human chondrocytes. The gene can be used
XX to study chondrite differentiation and degeneration mechanism to develop
XX gene therapy for deformative arthritis and chronic rheumatoid arthritis.
XX This technique may also be useful in therapy of diseases relating to
XX CAMP particularly induced by DC1 mRNA in other cell differentiation.
XX The gene is expressed specifically in differentiated human chondrocytes.
XX
XX Sequence 2948 BP; 762 A; 753 C; 716 G; 717 T; 0 other;

alignment_scores:
    Quality: 102.00      Length: 267
    Ratio: 0.761        Gaps: 19
    Percent Similarity: 50.187    Percent Identity: 28.464

alignment_block:
US-09-389-000-2 x X18940 ..

Align seg 1/1 to: X18940 from: 1 to: 2948

74 ProSerAsnThrValLysThrGluAsnAlaThrGlyProGluGluLeuG1 90
   |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
200 CCGGCGCATGGAGCGGATCCCGCGCGCAACACCCCGC..... 241
   |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
90  YLeuProLeuGlnArgSerTyrSer.GluHis.....LeuGlyTyr 103
   |||||  |||  ::::  ::::  ::::  ::::  ::::  ::::  |||
242 .CTGCGTCCCGCAAGACACCGGAGCTGAGCAGCAGACCTACAGAGATG 290
   ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  |||
104 PheProThrAspLeuPheAlaCysSerGluSerLeuArgAsnGlyAsnG1 120
   ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  |||
291 TACCGTCCCGCAATGTATCAAGTGTACAAGACAGCG.....GG 331
   ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  |||
120 YLeuGluLeuAsnAlaSerLeuSerGluPheGluLysAsnLysLysIle 137
   |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
332 AATAAAGGGGAGGAGACAGACAGAGAGACCTACAA..... 368
   |||  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
137 erLeuLeuHisSerSerLysGluLysLeuArgArgGluArgLleLysTyr 153
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
369 .TTGGCGGACACGGGCTCATCGAATAAGAGAGCTGACCGGATTAAAGAG 416
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
154 CysCysGluGluGlnLeuArgThrLeuLeuPro...TyrValLys..... 166

```

```

|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
417 TGCATGCCCGCCAGCTGAAGATCTCTACCCGAAACATCTCACTTACAAAC 466
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
167 ...GlyArgLysAsnAspAlaAlaSerValLeuGluAlaLeuThrValAsp 182
   |||  ::::  ::::  |||  |||  |||  |||  |||  |||  ::::
467 TTTGGGTCACTTG...GAAAGACAGTGGTCTTGAACCTTACCTTGAAGC 513
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
182 yValLysTyrIleArgGluLysIleSerProAlaValMetAlaGlnIle 198
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
514 ATGTGAAGCACATAACAACCTAATTGATCAG.....CAGCACAGAGAA 557
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
199 ThrGluAlaLeuGlnSerAsnMetArg..... 207
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
558 ATCATTTGCCCTGCAGAGTGTGTACAGCTGTGAGCTGCAGGAGAGAAA 607
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
208 .....PheCysLysLysGlnGlnThrProIleG 217
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
608 TGTGAAACAGGTCAAGAGATGTCTGCTCAGGTTTCCAGACATGTGCC 657
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
217 LLeuSerLeuProGlyThrValMetAlaGlnArgLysAsnSer..... 231
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
658 GGAGAGTGCTT.....CAGTATCTGGCCAGACAGAGAACATCTGGGAC 701
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
232 ..ValMetSer.ThrTyrSerProGluArgGlyLeuGlnPheLeuThra 247
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
702 CTGAAGTCTTGCAGCTGTGTCACCC.....ACCT 730
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
247 snThrCysTrpAsn.GlyCysSerThrProAspAlaGlnSerLeuAs 263
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
731 CCACCGGGGTGCTCGGAGCTGCT.....GCAGGGTGTCTCTTCCA 771
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
263 pGluAlaValArgValProSerSerSerAlaSerGluAsnAlaIleGly 280
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
772 GGAAGCATCAGA...CCAGCTCCCAAGTAT.....GGAC 806
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
280 sProTyrLysThrHisIleSerSerAlaAlaLeuSerLeuAsnSerLeu 296
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
807 TTCAAGCAAAAACCCAGCTCTCCGCCAAGGTTTCGG..... 844
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
297 HisThrValArgTyrTyrSerLysValThrProSerTyrAspAla 311
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
845 .....AGTCTCTGGGAAAACCTGCGTCCAGTATCTCAGCGC 880

```



```

US-09-389-000-2 x US-08-320-559-27
Align seg 1/1 to: US-08-320-559-27 from: 1 to: 9370

25 LysLysLysAsnThrHisSerLeuPheValPheIleIleProGluAsn 41
   ::::::::::::::::::::
2888 AGAGGAAATCAAAATCACAGTCTTCACTTCATCCCAAGAA 2937
   ::::::::::::::::::::
41 he.....LysGlyCysIleSerGlyHisGlyMetAspIleAla 54
   ::::::::::::::::::::
2938 TCTTCAAAACAA.....GCCCTC 2957
   ::::::::::::::::::::
55 ThrGluProLeuThrMetGluLysMetSerAsn.....65
   ::::::::::::::::::::
2958 CAGGCCCTCTCACAGTCTTCAAAAGAAATGCTCCCGCCAGCCG 3007
   ::::::::::::::::::::
66 .....ValVal. 67
   ::::::::::::::::::::
3008 TGTCTCTGCTCTCCAGAACCCAGCCAGCTGCACTTAAGAGTCAGG 3057
   ::::::::::::::::::::
68 LysTyrTrpThrThrCysProSerAsnThrValLysThrGluAsnAla 84
   ::::::::::::::::::::
3058 CGGGAAGCACACCTGTGGCCAGAGCCTCCCAAAAGTCCAGCAGTAC 3107
   ::::::::::::::::::::
84 r.....GlyProGluGluLeuGlyLeuProLeuGln.....94
   ::::::::::::::::::::
3108 CAGAGCAACACCAAAAGACTCTTCCATTCACAGCAGAGAAGTAGAG 3157
   ::::::::::::::::::::
95 .....ArgSerTyrSerGluHisLeuGlyTyr.....103
   ::::::::::::::::::::
3158 GGAAGGGCTCCAGAAAGCTCTCGAGCACAGAGGTCTTCGGAGATAC 3207
   ::::::::::::::::::::
104 .....PheProThrAspLeuPheAlaCysSerGluSerLeuArgAs 117
   ::::::::::::::::::::
3208 GCAAACTCTTCCAGTG.....CCTTCTTGGCAAA 3239
   ::::::::::::::::::::
117 nGlyAsnGlyLeuGluLeuAsnAlaSerLeuSerGluPheGluLysAsn 134
   ::::::::::::::::::::
3240 TGTAACTCT.....AAACAGGGAAGCCTCAAGTAGATTGACA 3280
   ::::::::::::::::::::
134 yLysLysIleSerLeuLeuHis.....SerSerLys 143
   ::::::::::::::::::::
3281 AACACACAGCACCTTCACATGAGGAGGACAAAAGATGAGAGCAAA 3330
   ::::::::::::::::::::
144 GluLysLeuArgArgGluArgIleLysTyrCysGluGluLeuArg 160
   ::::::::::::::::::::
3331 GCAGAGTTATACGACAGGAGGTGGAAGGCTTTAAGTACCTGGAAC 3380
   ::::::::::::::::::::
160 rLeuLeuProTyrValLys.....GlyA 168
   ::::::::::::::::::::
3381 CGTCTTGTCTTCATGATGCGGAATGCCACAGAGCTGAAGCCACT 3430
   ::::::::::::::::::::
168 rGlyAsnAspAlaAlaSerValLeuGluAlaThrValAspTyrVal 184
   ::::::::::::::::::::
3431 CATCAAGTACGCTTACTCTGTCTACTGAAACGTGATGATCTCAATA 3480
   ::::::::::::::::::::
185 TyrIle.....ArgG 188
   ::::::::::::::::::::
3481 TTCATTAAGTCTTAATAATCCTTCAGATGCCACAGCCCAACACAA 3530
   ::::::::::::::::::::
188 uLysIleSerProAlaValMetAlaGlnIleThrGluAlaLeuGln 205
   ::::::::::::::::::::
3531 GAAATATATTGCTGTTTATGCATGCGTGCACAGTCCATTTGACATG 3580
   ::::::::::::::::::::
205 smetLArgPheCysLysGlnGlnThrProIleGluLeuSerLeuPro 221
   ::::::::::::::::::::
3581 CGATGTTTCTGTTGTAAGAAAGC.....ATAGCATTAAGATATCT 3621
   ::::::::::::::::::::
222 GlyThrValMetAlaGlnArgLysAsnSerValMetSerThrTyrSer 238
   ::::::::::::::::::::
3622 CGTACTCTTAATAAACACTTCGAGAGTCTTCCAAAGTCCAGCCAGC 3671
   ::::::::::::::::::::
238 oGluArgGlyLeuGlnPheLeuThrAsnThrCysTrpAsnGlyCysSer 255

```

```

3672 T.....TCTCATGCAATGCAGACAGCA 3697
   ::::::::::::::::::::
255 hrProAspAlaGluSerSerLeuAspGluAlaValArgValProSer 271
   ::::::::::::::::::::
3698 CACCATCTCCCTTCTCCCAAGCCCTTCTCCGCGCAGCTCCGTAGGTC 3747
   ::::::::::::::::::::
272 SerAlaSerGluAsnAlaIleGlyAspProTyrLysThrHisIleSer 288
   ::::::::::::::::::::
3748 CAGTCAAGTGTGCGAGTGTGGGAGC.....AGTGG 3779
   ::::::::::::::::::::
288 rAlaAlaLeuSerLeuAsnSerLeuHisThrValArgTyrTyrSerLys 305
   ::::::::::::::::::::
3780 GGTGGCTGCACATACAGACAGCCAGTCACATCCAG.....ATA 3820
   ::::::::::::::::::::
305 alThrProSerTyrAspAlaThrAlaValThrAsnGlnAsnIleSer 321
   ::::::::::::::::::::
3821 TGACATCTTCTCTA.....GTCAACATC 3843
   ::::::::::::::::::::
322 HisLeuProSerAlaMetProVal.....330
   ::::::::::::::::::::
3844 ACAT...CCCATGTTCTTACCGCCTTTCACCTTGGGAAAGCCGAGCC 3890
   ::::::::::::::::::::
331 ..SerSerPheSerLeuGlyThrAlaLeuLeuGlyTyrPalaArgAla 346
   ::::::::::::::::::::
3891 CCTCACAGGAAGAAATAAAGATCTTCTGCTGGCTCAGCACAATGTGT 3940
   ::::::::::::::::::::
347 LeuHisIlePro.....ThrValCysAsnSerPheG 357
   ::::::::::::::::::::
3941 GCACCTGGCCCTCAACAGCAAGTTGTGGACCTGTGCACATACACA 3990
   ::::::::::::::::::::
357 yArgIleLysSerThr 362
   ::::::::::::::::::::
3991 CAGGCTTTCAGCAGC 4006
   ::::::::::::::::::::
seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-545-860D-27
seq_documentation_block:
: Sequence 27 Application US/08545860D
: Patent No. 6040140
: GENERAL INFORMATION:
: APPLICANT: Croce, Carlo
: APPLICANT: Canaan, Eli
: TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
: TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
: NUMBER OF SEQUENCES: 94
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
: ADDRESSEE: No. 6040140 is
: STREET: One Liberty Place, 46th floor
: CITY: Philadelphia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/545,860D
: FILING DATE: 07-MAR-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/04496
: FILING DATE: 22-APR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/10930
: FILING DATE: 09-DEC-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/327,392
: FILING DATE: 19-OCT-1994

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/320,559
FILING DATE: 11-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,443
FILING DATE: 14-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,094
FILING DATE: 30-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/888,839
FILING DATE: 27-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/805,093
FILING DATE: 11-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Deluca Esq., Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TUV-1262
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 9370 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 469..4032
US-08-545-860D-27

alignment_scores:
Quality: 113.00 Length: 424
Ratio: 0.589 Gaps: 22
Percent similarity: 45.283 Percent identity: 21.462

alignment_block:
US-09-389-000-2 x US-08-545-860D-27 ..
Align seg 1/1 to: US-08-545-860D-27 from: 1 to: 9370
25 LysLysLysAsnThrHisSer LeuPheValPheIleIleProGluAsn 41
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2888 AGAAGAAATCAATCAACAGTCATCTTCACTTCACTCCCAAGAA 2937
41 he.....LysGlyCysIleSerGlyHisGlyMetAspIleAlaLeu 54
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2938 TCTCTAAACAAA.....GCCCTC 2957
55 ThrGluProLeuThrMetGluLysMetSerAsn.....ValVal. 67
2958 CAGGCCCTCTCAAGCTCTCAAGGAAGATGCTCCCCCGCCACCCG 3007
66 .....ValVal. 67
3008 TGTCTCTGTCCTCCAGAGCCAGCCAGCCCTGCACTTAAGAGGTAAAG 3057
68 LysTyrTrpThrThcCysProSerAsnThrValLysThrGluAsnAla 84
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3058 CCGGAGACGAGACACCTGTGGCCAGCAGCCCTCCCAAAAGTGCAGCAGTAC 3107
84 r.....GlyProGluLeuGlyLeuProLeuGln..... 94
3108 CAAGAGCAACCAAGACTCTTCCATTCCTCCAGACGAGAGAGTAGAGG 3157
95 .....ArgSerTyrSerGluHisLeuGlyTyr..... 103
3158 GGAAGGGCTCCAGAAAGCTCTCGAGACCAAGAGGTCTCCGAGAGTACT 3207
```

```

104 .....PheProThrAspLeuPheAlaCysSerSerGluSerLeuArgAs 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3208 GCAAAATCCTTTCCAGTG.....CCTTTTGGCCAA 3239
117 ngLysnGlyLeuGluLeuAsnAlaSerLeuSerGluPheGlyLysAsn 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3240 TGGTAACTCT.....AAACGAGGAGAGCTCAATGAAGTTTGACA 3280
134 yLysIleSerLeuLeuHis.....SerSerLys 143
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3281 AACAAACAGACAGACTTCACATGAGGAGCGCAAAAGATGAGCAGAAA 3330
144 GluLysLeuArgArgGluArgIleLysTyrCysGlyGluGlnLeuArg 160
:::||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3331 GCAGAGTTAAATGACGACGAGGAGGTTGAAGGCTTTAAGTACTCGAAGC 3380
160 rLeuLeuProTyrValLys.....GlyA 168
:::||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3381 GCTCTTGCTTCATTTAGTGGGATTCGACAGATGCGCAAGCTGAAACCACT 3430
168 rGlyAsnAspAlaAlaSerValLeuGluAlaThrValAspTyrValLys 184
:::||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3431 CATCAAGTCAGCTTACTCTGCTCTCAGAAACTGTGATCTCATTA 3480
185 TyrIle.....ArgG 188
:::||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3481 TTCATATATGTCATTAATTCCTTCAGATGCGCACAGCGCCCAACAGAA 3530
188 uLysIleSerProAlaValMetAlaGlnIleThrGluAlaLeuGlnSer 205
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3531 GAATAATATTGCTGTGTTATGATCGGTGCGCAGTCCATTGGAACAATGG 3580
205 smMetArgPheCysLysGlyGlnGlnThrProIleGluLeuSerLeuPro 221
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3581 CGATCTTCGTTGTAAATAAGAC.....ATAGCAATAAGATATCT 3621
222 GlyThrValMetAlaGlnArgGluAsnSerValMetSerThrTyrSerPr 238
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3622 CGTACTCTTAATTAACACTTCGAGAGTTCTCCAAAGTCCGACGAGCAC 3671
238 oGluArgGlyLeuGlnPheLeuThrAsnThrCysTrpAsnGlyCysSer 255
3672 T.....TCTCATGTCATTGCAAGCACAGGCA 3697
255 hrProAspAlaGluSerSerLeuAspGluAlaValArgValProSerSer 271
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3698 CACCAATCCCTCTTCCCAATGCTCTGCGCAGAGTCCGATAGGATCC 3747
272 SerAlaSerGluAsnAlaIleGlyAspProTyrLysThrHisIleSerSe 288
:::||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3748 CAGTCAAGTCTGCGCAGTGTGGGAGC.....AGTGG 3779
288 rAlaAlaLeuSerLeuAsnSerLeuHisThrValArgTyrSerLysSV 305
:::||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3780 GGTGGCTGCACATATCAGACCCAGTCACATCCAG.....ATA 3820
305 alThrProSerTyrAspAlaThrAlaValThrAsnGlnAsnIleSerIle 321
:::||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3821 TGACATCTTCTAT.....GTCAACATC 3843
322 .HisLeuProSerAlaMetProProVal..... 330
|||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3844 ACAT...CCCATGTTCTTACCGCTTGAACCTTTGGAGACGGCCGAGGC 3890
331 ..SerSerPheSerLeuGlyThrAlaLeuLeuGlyTyrAlaArgArgAla 346
|||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3891 CCTCAGAGAGAAATTAAGATTTCTTGTCTGGCTCAGCAAAATGTGT 3940
347 LeuHisIlePro.....ThrValCysAsnSerPheG 357
|||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3941 GCACCTTGCCCTCAACAGACAGTTGTGTGACCTGTGCACATATACAGA 3990
357 yArgIleLysSerThr 362
```

```

seq_name: /cgr2_6/ptcddata/1/ina/PCUTUS_COMB.seq:PCT-US94-04496-27
||||: |||:
3991 CAGGCTTTTCAGACAC 4006

seq_documentation_block:
: sequence 27, Application PC/TUS9404496
: GENERAL INFORMATION:
: APPLICANT: Croce, Carlo
: APPLICANT: Canaan, Eli
: TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
: TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
: NUMBER OF INVENTION: Resulting from Chromosome Abnormalities in the All-1-1
: NUMBER OF SEQUENCES: 86
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
: ADDRESSEE: Norris
: STREET: One Liberty Place, 46th floor
: CITY: Philadelphia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/04496
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: DeLuca Esq., Mark
: REGISTRATION NUMBER: 33,229
: REFERENCE/DOCKET NUMBER: Tju-1242
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 568-3439
: TELEFAX: (215) 568-3100
: INFORMATION FOR SEQ ID NO: 27:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9370 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 469..4032
PCT-US94-04496-27

alignment_scores:
Quality: 113.00 Length: 424
Ratio: 0.588 Gaps: 22
Percent Similarity: 45.283 Percent Identity: 21.462

alignment_block:
US-09-389-000-2 x PCT-US94-04496-27 ..
Align seg 1/1 to: PCT-US94-04496-27 from: 1 to: 9370

25 LysLysLysAsnThrHisSer.LeupheValPheIleIleProGluAsp 11
:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2888 AGAAGAAATCAATACAGATCATCTTCATCTCATCTCCCAAGAA 2937
41 he.....LysGlyCysIleSerGlyHisGlyMetAspIleValLeu 54
:||||:||||:||||:||||:||||:||||:||||:||||:
2938 TCTCTTAAGACAA.....GCCCTC 2957
55 ThrGluProLeuThrMetGluLysMetSerAsp..... 65
2958 CAGGCTTCCTCAGTCATGCTCTCAAGAGAAAGAAATGCTCCCGCCGCCACCG 3007

```

66ValVal. 67

3008 TGTCTCGTCCTCCAGAAAGCCAGCAAGCCTGCACCTTAAGAGGTCAAG 3057

68 LysTyrTrpThrThrCysProSerAsnThrValLysThrGluAsnAlaThr 84

3058 CGGGAAGCAGACACTGTGGCCAGGAGACCTCCCAAAAGTGCACAGCATAC 31070

84 r.....GlyProGluGluLeuGlyLeuProLeuGln..... 94

3108 CAAGAGCAACCAACAAAGACTCTTCATTCGCCAAGAGGAAAGAGTAGAG 3157

95ArgSerYserGluHisLeuGlyTyr..... 103

3158 GGAAGGCTCCAGAAAGCTCTCTCGAGCACAAGGCTTCTCCGAGATACT 3207

104PheProThrAspLeuPheAlaCysSerGluSerLeuArgAs 117

3208 GCAAAATCCTTTCCAGT.....CCTCTTTGGCAAA 3239

117 nGlyAsnGlyLeuGluLeuAsnAlaSerLeuSerGluPheGluLysAsnL 134

3240 TGGTAACCT.....AAACGAGGAAAGCTCAAGTAGAAGTTTGCA 3280

134 ValLysLeuSerLeuLeuHis.....SerSerLys 143

3281 AACCAACAGCAACACTTACATGAGGAGGCAAAAAGATACACAGAAA 3330

144 GluLysLeuArgArgGluAlaArgIleLysTyrCysCysGluGluLeuArgTh 160

3331 GCAAGATTAATGACGACGAGGAGGTTGGAAAGCTTTAAGTACTGGAAGC 3380

160 rLeuLeuProTyrValLys.....GlyA 168

3381 CCGTCCTGCTTCATCATGATGGCGGAATGCCACAGAGTCTGAAGCCAGT 3430

168 rGlyAsnAspAlaAlaSerValLeuGluAlaThrValAspTyrValLys 184

3431 CATCCAGACGACGCTACTGCTGCTACTAGAAACTGTGATCTCATTTAA 3480

185 TyrIle.....ArgG 188

3481 TTCATTAATGACTTAATAATCCTTCACATGCCACAGGCCAACACAAAGA 3530

188 uLysIleSerProAlaValMetAlaGlnIleThrGluAlaLeuGlnSerA 205

3531 GAAATATATTGCTGTTTATGATCAGCGCTGGCAGCTCATTTTGAACATGC 3580

205 smMetArgPheCysLysLysGlnGlnThrProIleGluLeuSerLeuPro 221

3581 CGATCTTTCGTTGTAAAGAGC.....ATAGCAATTAATTAATTCT 3621

222 GlyThrValMetAlaGlnArgGluAsnSerValMetSerThrTyrSerPr 238

3622 CGTACTCTTAATAAACACTTGGAGAGTCTTCCAAAGTCCGCCAGGCACC 3671

238 oGluArgGlyLeuGlnPheLeuThrAsnThrCysTrpAsnGlyCysSer 255

3672 T.....TTCATGCGATGCAAGACACAGCA 3697

255 hrProAspAlaGluSerGluSerLeuAspGluAlaValAlaGlyValProSer 271

3698 CACCATCCCTCTTCCCAAGCTTCTCTCCAGCTCCGTGGGGTCC 3747

272 SerAlaSerGluAsnAlaIleGlyAspProTyrLysThrHisIleSerSe 288

3748 CAGTAAAGTCTGGGAGTGGGGAGC.....AGTGG 3779

288 rAlaAlaLeuSerLeuAsnSerLeuHisAlaThrValArgTyrTyrSerLys 305

3780 GGTGGCTGCACATACACACCCAGTCACCATCAG.....ATA 3820

305 aThrProSerTyrAspAlaThrAlaValThrAsnGlnAsnIleSerIle 321

```

      :||| |||||
3821 TGACATCTTCCTAT.....GTCACCATC 3843
3322 ..HisLeuProSerAlaMetProVal..... 330
3844 ACAT...CCCATGTTCTTACCGCTTGACCTTGGACAGCCCAAGC 3890
331 ..SerSerPheSerLeuGlyThrAlaLeuLeuGlyTrpAlaArgArgAla 346
3891 CCTCAGAGAGAGAAATAAAGAAATCTTGTGCTCGCAGCAGCAAAATGTGT 3940
347 LeuHisLeuPro.....ThrValCysAsnSerPheG1 357
3941 GCACCTTGCCCTCAACAGCAGCTTGTGTGACCTGTGCACTATACAGCA 3990
357 ValGlyLeuSerThr 362
3991 CAGGCTTTCAGCAGC 4006

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-320-559-25

seq_documentation_block:
: Sequence 25, Application US/08320559
: Patent No. 5633135
: GENERAL INFORMATION:
: APPLICANT: Croce, Carlo
: APPLICANT: Canaan, Eli
: TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
: TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
: TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
: NUMBER OF INVENTION: All-1 Region
: CORRESPONDENCE ADDRESSES:
: ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5633135rls
: STREET: One Liberty Place - 46th Floor
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: PC-DOS
: SOFTWARE: WORDPERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/320,559
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/062,443
: FILING DATE: 14 MAY 1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/971,094
: FILING DATE: 30-OCT-92
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/888,830
: FILING DATE: 27-MAY-92
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/805,093
: FILING DATE: 11-DEC-91
: ATTORNEY/AGENT INFORMATION:
: NAME: Deluca, Mark
: REGISTRATION NUMBER: 33,229
: REFERENCE/DOCKET NUMBER: TNU-0855
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 568-3100
: TELEFAX: (215) 568-3439
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9391 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
```

```

: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 421..4053
US-08-320-559-25

Alignment scores:
  Quality: 113.00      Length: 424
  Ratio: 0.589
  Percent Similarity: 45.283      Percent Identity: 21.462

Alignment_block:
US-09-389-000-2 x US-08-320-559-25 ..

Align seg 1/1 to: US-08-320-559-25 from: 1 to: 9391

25 LysLysLysAsnThrHisSerLeuPheValPheIleIleProGluAsn 41
: |||
: |||
2909 AGAAGGAAATCAATATCATCATCTTCACTTCACTCCACAAAGAA 2958
41 he.....LysGlyCysIleSerGlyHisGlyMetAspIleAlaLeu 54
: |||
: |||
2959 TCTCTCAAAACAA.....GCCCTC 2978
55 ThrGluProLeuThrMetGluLysMetSerAsn..... 65
2979 CAGGCCCTCTCAGAGCTCTCAAGAAAGAAATGTCTCCCGCCAGCCG 3028
66 .....ValVal. 67
3029 TGTCTCTGTCCTCCAGAAAGCCAGCCAGCTGCACCTTAAGAGTCAAG 3078
68 LysTyrTrpThrThrCysProSerAsnThrValLysThrGluAsnAlaThr 84
: |||
: |||
3079 CCGGAGACAGACACCTGTGTGCGCAGAGCCCTCCCAAAAGCCAGCAGTAC 3128
84 r.....GlyProGluGluLeuGlyLeuProLeuGln..... 94
: |||
: |||
3129 CAAGACACACCAAAAGACTTTCATCTCCCAAGAGAGAGAGTAGAGG 3178
95 .....ArgSerTyrSerGluHisLeuGlyTyr..... 103
3179 GGAAGGCTCCAGAGAGCTCTCGACACAGAGGTTCTCCGAGATACT 3228
104 .....PheProThrAspLeuPheAlaCysSerGluSerLeuArgAs 117
3229 GCAATCCTTTCAGAGT.....CCTCTTGGCAAA 3260
117 nGluAsnGlyLeuGluLeuAsnAlaSerLeuSerGluPheGluLysAsn 134
3261 TGGTAACTCT.....AAACAGGAAAGCTCAAGTAGAATTGACA 3301
134 yLysIleSerLeuLeuHis.....SerSerLys 143
: |||
: |||
3302 AACCAACAGCAGACCTTCACATGAGGAGGCAAAAAGATGACGAA 3351
144 GluLysLeuArgGluArgIleLysTyrCysGlyGluLeuArgThr 160
3352 GCAGAGTTAATGACGACAGGAGGTGGAAGGCTTTAATGACTGGAAC 3401
160 rLeuLeuProTyrValLys.....GlyA 168
3402 CGCTGTGCTCATATGAGTGGGAATTCACAGAGTGTGAAGGCACT 3451
168 rLysAsnAspAlaAlaSerValLeuGluAlaThrValAspTyrValLys 184
3452 CATCCAAGTACACTTCTGTCTACTCAGAAACTGTAGATCTCATTA 3501
185 TyrIle.....ArgS1 188
3502 TTCATTAATGTCATTAATATCTTTCAGATGCCACAGGCCCAACACAA 3551
```

```

188 uLySleSerProAlaValMetAlaGlnIleThrGluAlaLeuGlnSera 205
||||| : : : : : ||| : : :
3552 GAAATATTTCTCTTTATGATCGCGTTCACATGTTTGAACATGG 3601
205 smMetArgPheCysLysGlnInThrProIleGluLeuSerLeuPro 221
||||| : : : : : ||| : : :
3602 CGATGTTTCGTTAAAGAACAC.....ATAGCAATTAAGATATCT 3642
222 GlyThrValMetAlaGlnArgGluAsnSerValMetSerThrSerPr 238
||||| : : : : : ||| : : :
3643 CGTACTCTTAATAACACTTCGAGAGTCTTCCAAAGTCCGACGACAC 3692
238 ogLArgGlyLeuGlnPheLeuThrAsnThrCysTrpAsnGlyCysSer 255
| : : : : | : : : : | : : : : | : : : : |
3693 T.....TCTCCATGCATTCGACACAGCACA 3718
255 hrProAspAlaGlnSerSerLeuAspGluAlaValArgValProSerSer 271
||||| : : : : : ||| : : :
3719 CACCATCCCTCTTCCCATGCTTCTCTGCGACGCTCCGAGGCTCC 3768
272 SerIaSerGluAsnAlaIleGlyAspProTyrIleThrHisSerSe 288
| : : : : | : : : : | : : : : | : : : : |
3769 CAGCAAGTCTGCGAGTGGGAGC.....AGTGG 3800
288 rAlaAlaLeuSerLeuAsnSerLeuHisThrValArgTyrSerLysV 305
| : : : : | : : : : | : : : : | : : : : |
3801 GGTCGCTGCCATCTACAGACCCGACATCAG.....ATA 3841
305 alThrProSerTyrAspAlaThrAlaValThrAsnGlnAsnIleSerLe 321
| : : : : | : : : : | : : : : | : : : : |
3842 TGACATCTCTCTAT.....GTCACCATC 3864
322 HisLeuProSerAlaMetProProVal..... 330
| : : : : | : : : : | : : : : | : : : : |
3865 ACAT...CCATGTTCTTACCGCTTGGACCTTGGACAGCCGAGGC 3911
331 SerSerPheSerLeuGlyThrAlaLeuLeuGlyTyrPalaArgAla 346
| : : : : | : : : : | : : : : | : : : : |
3912 CCTCAGCAGAGAAATTAAGATCTTTCGCGCTCAGCAATATGCT 3961
347 LeuHisIlePro.....ThrValCysAsnSerPheG 357
| : : : : | : : : : | : : : : | : : : : |
3962 GCACCTTGCCCTCAACAGCAGTTGTGTGACCTGTGCACTATACAGA 4011
357 YArgIleLysSerThr 362
| : : : : | : : : : | : : : : | : : : : |
4012 CAGGCTTTCAGCAGC 4027

seq_name: /cgn2_6/plodata/1/lna/6_COMB.seq:us-08-545-860D-25
seq_documentation_block:
: Sequence 25, Application US/08545860D
: Patent No. 6040140
: GENERAL INFORMATION:
: APPLICANT: Croce, Carlo
: APPLICANT: Canani, Eli
: TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
: TITLE OF INVENTION: for detection and treatment of Acute Leukemias
: TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1 Region
: NUMBER OF SEQUENCES: 94
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &
: STREET: One Liberty Place, 46th floor
: CITY: Philadelphia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/545,860D
: FILING DATE: 07-MAR-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/04496
: FILING DATE: 22-APR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/10930
: FILING DATE: 09-DEC-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/327,392
: FILING DATE: 19-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/320,559
: FILING DATE: 11-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/062,443
: FILING DATE: 14-MAY-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/971,094
: FILING DATE: 30-OCT-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/888,839
: FILING DATE: 27-MAY-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/805,093
: FILING DATE: 11-DEC-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Deluca Esq., Mark
: REGISTRATION NUMBER: 33,229
: REFERENCE/DOCKET NUMBER: T2U-1262
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 568-3100
: TELEFAX: (215) 568-3439
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9391 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 421..4053
: US-08-545-860D-25

alignment_scores:
: Quality: 113.00 Length: 424
: Ratio: 0.589 Gaps: 22
: Percent Similarity: 45.283 Percent Identity: 21.462

alignment_block:
US-09-389-000-2 x US-08-545-860D-25 ..
Align seg 1/1 to: US-08-545-860D-25 from: 1 to: 9391

25 LysLysLysAsnThrHisSer.LeuPheValPheIleIleProGluAsn 41
: : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : :
2909 AGAAGCAATCAATCAACAGTCTTCACTTCACTCCCAAGAA 2958
41 he.....LysGlyCysIleSerGlyHisGlyMetAspIleAlaLeu 54
: : : : : : : : : : : : : : : : : : : : : : : : : :
2959 TCTTCTAAACAAA.....GCCCTC 2978
55 ThrGluProLeuThrMetGluLysMetSerAsn..... 65
: : : : : : : : : : : : : : : : : : : : : : : : : :
2979 CAGGCCCTCTCTACACATCCCAAGAAAGAAATGCTCCCGCACCG 3028
66 .....ValVal. 67
3029 TGCTCTGCTCTCCAGAACGACCAAGCCTGCACTTAAGAGTCAAG 3078

```


68	LYSTIPTPTHTHCYSPROSERASNTHTVALYSTHTGLUASALATH	84
3079	CGGGAAAGCAGACACCTGTGTGCCAGGACCTCCCAAAAGTCCAGCAGTAC	3128
84	r.....GIYPROGLUGLULEUGLYLEUPROEUGLN.....	94
3129	CAGAGCAACCAACAAAGACTCTTCCTATTCGCCAAGCAGAGAGATGAGG	3178
95ArgserTYRserGIUHisLeuGLYTYr.....	103
3179	GGAAAGGCTCCAGAAACCTCTCGGAGACACAAAGGTTCTTCGGAGATACT	3228
104PheProthrAspLeupheALacysserGIUserLeuArgAs	117
3229	GCAAAATCCTTTTCCAGTG.....CCTCTTTCACAA	3260
117	ngLYasnGLYLeuGLULeuAsnAlaSerLeuSerGIUpheGIUysAsnL	134
3261	TGGTAACTCT.....AAACCAGGGAAGCCTCAAGTGAAGTTTGACA	3301
134	ysysTLserLeuLeuHis.....SerserLYs	143
3302	AACAAACACAGACCTTCTACATGAGGAGCGCAAAAAGATGAAGCAGAA	3351
144	GLUYLeuArgArgGLUArgILEYTYrCYsGSLUGLULeuArgTH	160
3352	GCGAGATTATATAGCGACAGCGGTGGAAAGGCTTTTAAGTACCTGGAAC	3401
160	rLeuLeuProTYrValLYs.....GLY	168
3402	CGCTTGTCTCTTCACTGAGTGGGGAATGGCCAGAGTCTGAAACCACT	3451
168	rGLysAsnAspAlaAlaSerValLeuGLUAlaThrValAspTYrValYs	184
3452	CATCCAGACAGCTACTCTGTCTACTCGAAGACTGTAGATCTCATTTAA	3501
185	TYRTLE.....ArgI	188
3502	TTCATATGTCAATTAAATCCTTCTCAGATGCCACAGCGCCAAACAAAGA	3551
188	uysTLserProAlaValMetAlaGLInLehrGIUAlaLeuGlnserA	205
3552	GAAAAATTTTGGCTGTTTATGCATGGGTGGCCAGTCCATTGTGAACATGG	3601
205	smetArpPheCYsLYsGSLngInThrProILEuSerLeuPro	221
3602	CGATGTTCGTTGTATAAAAAGC.....ATAGCAATAAAGATATCT	3642
222	GLYTHValMetAlaGLInArgGLUAsnserValMetSerTHrTYrSerPr	238
3643	CGTACTCTTAATAAACACTTCGAGAACTTTCCAAAGTCGCCAGCGAAC	3692
238	oGIUAArgLYLeuGLInPheLeuThrAsnThrCYsTrpAsnGLYCYsSerT	255
3693	T.....TCCTCATGCACTTGCAGACACAGCA	3718
255	hrProAspAlaGLUserSerLeuAspGLUAlaValArgValProSerSer	271
3719	CACATACCCCTCTTTCGCCAAAGCCCTTCCCGCCAGCTCCGTAGGCTTC	3768
272	serAlaSerGIUAsnAlaILEgIAspProTYrLYsTHHisLEserSe	288
3769	CATCACTAGCTCGCGAGTGTGGGAGC.....AGTGG	3800
288	rAlaAlaLeuSerLeuAsnSerLeuHisThrValArgTYrTYrSerLYsV	305
3801	GGTGGCTCGACATATACGACACCCAGCTCACATCCAG.....AATA	3841
305	alhrProserTYrAspAlaThrAlaValThrAsnGLInAsnLEserILE	321
3842	TGACATCTTCTAT.....GTCAACATC	3864

```

332  HisLeuProSerAlaMetProPheVal..... 330
    ||| ||| ::|||:::
3865 ACAT...CCCATGTTCTTACCGCCCTTTGACCTTTGGGAAACAGCCGACGC 3911
331  ..SerSerPheSerLeuGlyThrAlaLeuLeuGlyTTrpAlaArgArgAla 346
    ||| :::::||||| |||:::
3912 CCTCAGCAGGAAGATATAAGATTCTTGGCTCGCGTCAGACACAAGATGT 3961
347  LeuHisLePro.....ThValCysAsnSerPheG1 357
    ||| |||:::
3962 GCACCTTGCCCTCAACAGCAGTTGGTGGACCTGTGCACATATACAGA 4011
4012 CAGGGGTTTTCAGCAGC 4027

seq_name: /cgn2_6/prodata/1/lna/PCTrus_comb.seq:PCT-US94-04496-25

seq_documentation_block:
: Sequence 25, Application PC/US9404496
: GENERAL INFORMATION:
: APPLICANT: Croce, Carlo
: TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
: TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
: NUMBER OF INVENTIONS: Resulting from Chromosome Abnormalities in the All-1
: NUMBER OF SEQUENCES: 86
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
: ADDRESSEE: Norris
: STREET: One Liberty Place, 46th floor
: CITY: Philadelphia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/04496
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Deluca Esq., Mark
: REGISTRATION NUMBER: 33,229
: REFERENCE/DOCKET NUMBER: TUD-1242
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 568-3100
: TELEFAX: (215) 568-3439
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9391 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 421..4053
: PCT-US94-04496-25

alignment_scores:
Quality: 113.00 Length: 424
Ratio: 0.588 Gaps: 22
Percent Similarity: 45.283 Percent Identity: 21.462

alignment_block:
US-09-389-000-2 x PCT-US94-04496-25 ..
Allgn seg 1/1 to: PCT-US94-04496-25 from: 1 to: 9391

```

```

255 hProxaplaIagIserSerleuaspoluIaValaIargValaProSerSer 276
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3719 CACATCCCTCTTCTCCCAATGCTCTTCTCTCCACCTCCGTAAGGCTC 3768
272 SerAlaSerGIuAsnaIaIleGIaaspProTylrYlSerThrHisIleSer 288
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3769 CAGTCAAGAGTCTGGCAGTGTGGGAGC.....AGTGG 3800
288 rAlaAlaIeSerleuAsnSerleuHisThrAlaArgTyrTyrSerIysV 305
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3801 GGGGGTGGCAGTATAGCAGCCAGTCACATTCAG.....ATA 3811
305 aAlrProSerTyrAlaPaIaThAlaValThrAsnGlnAsnIleSerIle 321
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3842 TGACATCTTCCTAT.....GTACATC 3864
322 .HisIeProSerAlaMetProProVal..... 330
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3865 ACAT...CCATGTTCTTACCGCTTGGCTTGGAGACAGCCGAGC 3911
331 ..SerSerPheSerIeAluGIYThAlaIeIuGIYTrPaIaArgAra 346
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3912 CCTCAGGAGAGAAATAAAGAAATTTCTTCTCGGCTCAGCAAAATGT 3961
347 LeuHisIlePro.....ThValCysAsnSerPheG1 357
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3962 GCACCTTGCCCTTCACAGACGATGTGTGGAGCTGGTGCACTATATACGA 4011
357 yArgIleYsserThr 362
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4012 CAGGGTTTCAGAGC 4027
seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-528-199-2
seq_documentation_block:
: Sequence 2, Application US/08528199
: Patent No. 5763228
: GENERAL INFORMATION:
: APPLICANT: KUBOTA, Michio
: APPLICANT: TSUSAKI, Keiji
: APPLICANT: SUGIMOTO, Toshiyuki
: TITLE OF INVENTION: RECOMBINANT ENZYME FOR CONVERTING
: TITLE OF INVENTION: MALTOSE INTO TREHALOSE
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BROWDY AND NEIMARK
: STREET: 419 Seventh Street, N.W., Suite 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/528,199
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/485,126
: FILING DATE: 14-SEP-1995
: APPLICATION NUMBER: US 08/485,126
: FILING DATE: 07-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 156399/1994
: FILING DATE: 16-JUN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: BROWDY, Roger L
: REGISTRATION NUMBER: 25,618
: REFERENCE/DOCKET NUMBER: KUBOTA-6A
: TELECOMMUNICATION INFORMATION:

```



```

REFERENCE/DOCKET NUMBER: KUBOTA-6A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1704
US-08-528-199-5

alignment_scores:
Quality: 109.50      Length: 351
Ratio: 0.730         Gaps: 14
Percent Similarity: 42.735      Percent Identity: 20.513

alignment_block:
US-09-389-000-2 x US-08-528-199-5 ..

Align seg 1/1 to: US-08-528-199-5 from: 1 to: 1704

44 CysIleSerGlyHisGlyMetAspIleAlaLeuThrGluPro..... 57
||| ||||| ||| :: :: ||| |||
680 TGCTCAAGCGGCGTGGCGCTTCGTCGACGACACTACCCGACGCGGCTG 729
58 LeuThrMetGluysMetSerAsnValValLysTyrTrpThrProCysP 74
::::: ||||| ::||| ::||| ::||| ::|||
730 CTGCTGACGAGCGCAACCACTGCGCCGACGACGCGTGGAGTACTTCGG 779
74 roSerAsnThrValLysThrGluAsnAlaThrGlyPro.GluGluLeuG1 90
::::: ||||| ::||| ::||| ::||| ::|||
780 GCCCGAGAGCGTGCAGCGACGCGCTGCGCGCCGACAGTCCATCAGG 829
90 yLeuProLeuGlnArgSerTyrSerGlnHisLeuGlyTyrPhePro.Thr 106
||||| ||||| ::||| ::||| ::|||
830 CCTTCCACTTCCCGTATGCGCGCATCTTCATGCGCGTGCAGCGCGAG 879
107 AspLeuPheAlaCysSerGluSerLeuArgAsnGlyLeuGluLe 123
::||| ||||| ::||| ::||| ::|||
880 TCGCGCTTCCGATCTCGAGATCATGAGCAAGACCGCGGATCCCGGA 929
123 uAsnAlaSerLeuSerGluPheGluLysAsnLysLysLysLeuLeu 140
::::: ||||| ::||| ::||| ::|||
930 GGGCTGCAGTGGGGCATCTTCGCGCACAACGACGAGTCACTCG 979
140 lSerSerLysGluLysLeuArgGluArgGlyLeuLysTyrCysGlu 156
::||| ::||| ::||| ::|||
980 AGATGTCACCGAGAGACCGC..... 1002
157 GlnLeuArgThrLeuLeuProTyrValLysGlyArgLysAsnAlaAl 173
||| ::||| ::|||
1003 .....GACTACATGTGGGCG..... 1017
173 aserValLeuGlnAlaThrValAspTyrValLysTyrIleArgLysI 190
::||| ::||| ::|||
1018 .....GAGTACGCCAAGACCCCGCATGAAG 1045
190 lSerProAlaValAlaMetAlaGlnIleThrGluAlaLeuGlnSerAsnMet 206
::||| ::||| ::|||
1046 CCAACATCGCATCCGCGCGCGCTCGCGCTCGTGCAGACAGACAGC 1095
207 ArgPheCysLysLysGlnGlnThrProIleLeuLeuSerLeuProGly 223
::||| ::||| ::|||
1096 AACCAAGATC...GAGCTGTTACCGCGCTGCTGCTGCTGCTGCTG 1140
223 rValMetAlaGlnArgGluAsnSerValMetSerThrTyrSerProGlu 240

```

```

1141 ..... ||| ::||| ::|||
240 rgGlyLeuGlnPheLeuThrAsnThrCysTrpAsnGlyCysSerThrPro 256
||| ::||| ::|||
1169 TCGGC.....ATGGCGACAAACATCTGGCTCGGTGACCCG..... 1203
257 AspAlaGluSerSerLeuAspGluAlaValArgValProSerSerAl 273
::::: ||||| ::||| ::|||
1204 .....GACGGCTGGTACGGCGATGACGGAGC 1232
273 aserGluAsnAlaIleGlyAspProTyrLysThrHisIleSerSerAla 290
::::: ||||| ::||| ::|||
1233 CCGCGACCGCAAGCTCGC..... 1251
290 lalSerLeuAsnSerLeuHisThrValArgTyrTyrSerLysValThr 306
::||| ::|||
1252 .....TTCTCGCGCGCCACG 1266
307 ProSerTyrAspAlaThrAlaValThrAsnGlnAsnIleSerIleHis 323
||| ::||| ::|||
1267 CCGCGC.....AAGCTGCACT 1283
323 uProSerAlaMetProProValSerSerPheSerLeu..... 335
||| ::||| ::|||
1284 GCGAGCATCCAGGACCGCGTCTACGGCTACAGAGCGTCAAGTCGAGG 1333
336 .....GlyThrAlaLeuLeuGlyTyrTrpAlaArgAlaLeu 347
::::: ||||| ::||| ::|||
1334 CGCAGCTGAGAACCCCTCCGCTGCTGCACGTCGACCGCGGATGATC 1383
348 HisIleProThrValCysAsnSerPheGlyArgIleLysSerThrCysLe 364
||||| ::||| ::|||
1384 CACATCCGCGCCGACGCGGACCTTCGG..... 1413
364 uLysPheThrLeuSerThrThrTyrTrpAlaGlnPheAspAsnLeuGly 380
||| ::||| ::|||
1414 .....CTGGCAGC.....TTCGAGAGACTCGC 1437

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-049-282B-1

seq_documentation_block:
; Sequence 1, Application US/08049282B
; Patent No. 5534660
; GENERAL INFORMATION:
; APPLICANT: CHUCK, George S.
; APPLICANT: DOONER, Hugo K.
; APPLICANT: COURTNEY-GUTTERSON, Neal
; APPLICANT: KELLER, Janis
; APPLICANT: NIJAR, Charanjit S.
; APPLICANT: RALSTON, Edward J.
; TITLE OF INVENTION: PH GENES AND THEIR USES
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourile and Crew
; STREET: Stewart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94105-1493
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/049,282B
; FILING DATE: 16-APR-1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 12176-34

```

```

? TELECOMMUNICATION INFORMATION
? TELEPHONE: (415) 543-9600
? TELEFAX: (415) 543-5043
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1149 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: CDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..1146
US-08-049-282B-1

```

```
alignment_scores:
  Quality: 108.50
  Ratio: 0.714
  Percent Similarity: 49.673
  Length: 306
  Gaps: 15
  Percent Identity: 23.529
```

```
alignment_block:
US-09-389-000-2 x US-08-049-282B-1 . .
```

Align seg 1/1 to: US-08-049-282B-1 from: 1 to: 1149

6 ValProSerSerLeuSerAlaGluLeuLeuAlaIleLysLeuLeu 22
 187 TTACCAAGCGCGCCCTAGTATGATCAATA..... 216
 22 gphegLYLysLYsAsnThrHs...SerLeuPheValPheIleP 38
 217TCAAGNAGAACACACATTTACTCTAAACAGTGCAGCATTC 259
 38 roGLuAsnPhelys.....Gly 43
 260 TTGAACACCTCTCAACCAAGCTCCAAATTTCTCTACCATTAAGGC 309
 44 CysIleSerGLYHisGLYMetaspIleAlaLeuThrGLuProLeuThrMe 60
 310 TGTATTTC..... 318
 60 tGLuYsMetSerAsnValLYsTYrTyrPheThrCysProSerAsnT 77
 319 .CAACACCAACCAATCTGCTTCACAGGTGGCCAGC...CCAGACCA 364
 77 hValLYsThrGLuAsnAlaThrGLYProGLuLeuGLYLeuProLeu 93
 365 CGCTTCTCCAGCCATTTCTTGACGGCGGGCCACCTCGGCGCAGTGGCTG 414
 94 GLuArgSerTYrSerGLuHisLeuGLYTYrPheProThrAspLeuPheAl 110
 415 CTCAAAAGCATACTATTCTCTGTTCCATTCTTCACACTAAATATCCAAAC 464
 110 acYSerGLu...SerLeuArgAsnGLYAsnGLYLeuGLuLeuAsnAlas 126
 465 TGCAGCTGAAGTTCTCCAAAGTCAAGCAGACCTACACACTGTGATTCTC 514
 126 er...LeuSerGLuPheGLuLYsAsnLYsIle..... 136
 515 CCACTGCATCTCGCTTCGAAAGGGGTAGTATTAACACAGAGAGACCT 564
 137 SerLeuLeuHisSerSerLYsGLuLYsLeuArgArgGLuArgIleLYsTY 153
 565 AGTGGAAACCATGTACTTGCTGCAAGACGGCGCTAGGAAGTCAACGA 614
 153 rCYsCYsGLuLeuAlaArgThrLeuLeuProTYrValLYsGLYArgLYsA 170
 615 ACGTTATATCATTTTGAGAGTCACTTCCTTCCTTTGT.....ACGAAA 658
 170 snAspAlaIleSerValLeuLeuAlaThrValAspTYrValLYsTYrIle 186
 659 TGAATTAAGCTTCATTTCTGCTGACACCATTAAGAAATATTCACAGCTTA 708

```

187 ArgGluTyrLysSerPro...AlavMetIacInlThrcguAate 2032
   |||::|||:::::::::: :::: |||::: ||||||||
709 CGTAGGAAGAATTCCAGATCTTGAAAGCGTAAGACCATCAGACGGAGGT.. 756
      uGlSerAsnMetArgPheCysLysLYSGinglnThrPrILeGIuleus 219
      ..... :::::::::::::::::::: ACCCTGCAGACA 769
      757 .....,.....
219 erLeuProGLYrThValMeLaclnaIrgrLuasnsErValmetsErThr 2355
   :: |||||||| :~::~ :::: :::::
770 AGCAATCACGGACTGTGAAGGTGTTCGACAGAACGGGT..... 807
236 TyrSsrProGUArvggLylenGlnPhelenuThraSnHrcystrPaengl 2522
     808 ..... MAGCGAATATAAATAGTATCGAAGAAAGTTGTGTGTGAGC 848
           ::||| :::::::::::::: ::::: :::::
232 Y.....CysserthrProaspalaGluSerSerleuaSpc 2644
    |-----::|::|::|::|::|::|::|::|::|::|::|::|::|
849 ACAGCAAAGATCACGCCGACTCTTCGCCCTCACACACATCAAA...GAGC 895
264 LuAlaValArgValProserSerSerlaaserGiuaenAIaleGIlyASP 2800
    || |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
896 AEAATGTCAAAGTGAAGACTATATACGAAGTATGACGACTGTGTCAG 945
281 .....ProtYrLYs 283
       |||::|::|::|
946 CTCAGCTGTCCATNCAA 963

seq_name:/cnj_2_6/prodatta/1/inia/5B_COMB.seq:US-08-537-715-1
seq_documentation_block:
; Sequence 1, Application us/08537715
; Patent No. 5910627
GENERAL INFORMATION:
APPLICANT: Chuck, George S.
APPLICANT: Dooner, Hugo K.
APPLICANT: Courtney-Guterson, Neal
APPLICANT: Keller, Janis
APPLICANT: Nijjar, Charand J.S.
APPLICANT: Ralston, Edward J.
TITLE OF INVENTION: PH Genes and Their Uses
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
City: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,715
FILING DATE: 16-FEB-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/049,282
FILING DATE: 16-Apr-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US94/04173
FILING DATE: 15-Apr-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 012176-003410US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEX: (415) 576-0300
INFORMATION FOR SEQ ID NO.: 1:
```

```

: SEQUENCE CHARACTERISTICS:
: LENGTH: 1149 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1149
: OTHER INFORMATION: /note= "ph gene cDNA"
US-08-537-715-1

```

```

alignment_scores:
  Quality: 108.50      Length: 306
  Ratio: 0.714        Gaps: 15
  Percent Similarity: 49.673   Percent Identity: 23.529

```

alignment_block:
US-09-389-000-2 x US-08-537-715-1 ..

Align seg 1/1 to: US-08-537-715-1 from: 1 to: 1149

```

6 ValProSerSerLeuSerAlaGluGluLeuGluAlaIleLysLeuIleAr 22
  ::::::::::: ||| :::::::::::
187 TTACGAGGGGCGCTAGTATGATGATTA..... 216
22 GpheLysLysLysAsnThrHis...SerLeuPheValPheIleIleP 38
  ::::::::::: ||| |||
217 .....TCACAGAAAGACACACATTACTCTCAACAGTGCAGCCATT 259
38 roGluAsnPhelys.....Gly 43
  |||:::::::::: |||
260 TTGAACACCTCTCAACCAAGCTCCAATTTCCTCCATCATGAGCC 309
44 CysIleSerGlyHisGlyMetAspIleAlaLeuThrGluProLeuThr 60
  ||| ||| |||
310 TGTATTCC..... 318
60 tGluLysMetSerAsnValAlaLysTyrThrPheThrCysProSerAsn 77
  ::::::::::: ||| :::::::::::
319 .CAACCAACCCCAATCTGCTTCACAGGCGGCCAGC...CCAGCCACA 364
77 hrValLysThrGluAsnAlaThrGlyProGluGluLeuGlyLeuPro 93
  ||| ||| |||
365 CCGTTCACAGCCATTCTTGACGGCGGCCACCTCCGGCCAGTGGCTG 414
94 GluArgSerTyrSerGluHisLeuGlyTyrPheProThrAspLeuPhe 110
  ::::::::::: ||| :::::::::::
415 CTCAAAACCATCTATCTCTGTCATTCTTCACACTAATAATACCAAC 464
110 acYserGlu...SerLeuArgAsnGlyAsnGlyLeuGluLeuAsnAla 126
  ::::::::::: ||| :::::::::::
465 TGCAGCTGAAGTTTCTCCAAAGTCAGTGCCTACCACTGTTGATTCT 514
126 er...LeuSerGluPheGluLysAsnLysLysIle..... 136
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
515 CCACATGCATCTGCTTGAAAGGCTAGATAACACAAAGAGCCCT 564
137 SerLeuLeuHisSerSerLysGluLysLeuArgArgGluArgIleLys 153
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
565 AGTGAAACCATGATCTGCTGACAGCGGCGTAGAAGAAACCTCAACGA 614
153 rCysCysGluGluLeuArgThrLeuLeuProTyrValLysGlyArgLys 170
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
615 ACGGTTTATCATTTGAGGTCACTGTGCTTTGTT.....ACGAAA 658
170 snAspAlaLaserValLeuGluAlaThrValAspTyrValLysTyrIle 186
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
659 TGGATTAAGCCCTCATCTCTGTGACACCATATGATATGCAAGCAGTA 708
197 ArgGluLysIleSerPro...AlaValMetAlaGlnIleThrGluAla 202
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

709 CETAAGAAAGTTACAGATCTTGAAAGCTAGAGCAATCAGAGGAGCT.. 756
202 uGlnSerAsnMetArgPheCysLysLysGlnGlnThrProIleGluLeu 219
757 .....ACGCTGCAGACAA 769
219 erLeuProGlyThrValMetAlaGlnArgGluAsnSerValMetSerThr 235
  ::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
770 AGGATACAGAGTACTCTGAAAGGCTTGCAAGAGAGGCT..... 807
236 TyrSerProGluArgGlyLeuGlnPheLeuThrAsnThrCysTrpAsnG 252
  ::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
808 .....AAGAGGAGATAGTACAGAGAGAGAGTGTGTGAGAG 848
252 Y.....CysSerThrProAspAlaGluSerSerLeuAspG 264
849 ACAGGCAAGATCAGCGATCTTCGCTCAACGACACATGAA...GAGG 895
264 LuAlaValArgValProSerSerSerAlaSerGluAsnAlaIleGlyAsp 280
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
896 AGATAGTCAGAGTACAGATATCATATTCAGAGAGTATGCTGTGAGAG 945
281 .....ProTyrLys 283
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
946 CTCAGGTGTCATACAAA 963

```

seq_name: /cgn2_6/prodata/1/ina/PCTUS_COMB.seq:PCT-US94-04173-1

```

seq_documentation_block:
: Sequence 1, Application PC/RUS9404173
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: PH GENES AND THEIR USES
: NUMBER OF SEQUENCES: 4
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/04173
: FILING DATE: 15-Apr-1994
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/049,282
: FILING DATE: 16-Apr-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Bastian, Kevin L.
: REGISTRATION NUMBER: 34,774
: REFERENCE/DOCKET NUMBER: 12176-34-1PC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 543-9600
: TELEFAX: (415) 543-5043
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1149 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1146
: PCT-US94-04173-1

```

```

alignment_scores:
  Quality: 108.50      Length: 306
  Ratio: 0.714        Gaps: 15
  Percent Similarity: 49.673   Percent Identity: 23.529

```

alignment_block:
US-09-389-000-2 x PCT-US94-04173-1 ..

Align seg 1/1 to: PCT-US94-04173-1 from: 1 to: 1149

```

6 ValProSerSerLeuSerAlaGluLeuGluAlaIleLeuIleLeu 22
   ::::::::::: ||| :::::::::::
187 TTACAGGCGCGCCGATGATGATGATTA..... 216
22 pPheGlyLysLysLysAsnThrHis...SerLeuPheValPheIleLeu 38
   ::::::::::: ||| :::::::::::
217 .....TCACAGAGAGACACACATTACTCTCAACAGAGTGCACATTC 259
38 roGluAsnPheLys.....Gly 43
   ||| :::::::::::
260 TTGAACACCTCTCAACACCAAGCTCCAAATTTCTCTACCAATATGGGC 309
44 CysIleSerGlyHisGlyMetAspIleAlaLeuThrGluProLeuThrMe 60
   ||| :::::::::::
310 TGTATTTCC..... 318
60 tGluLysMetSerAsnValAlaLysLysTyrTrpThrCysProSerAsn 77
   ::::::::::: ||| :::::::::::
319 .CAACAAACCCAAATCTGCCTTCACAGAGTGCCAGC...CCAGACACA 364
77 hValLysThrGluAsnAlaThrGlyProGluGluLeuLeuProLeu 93
   ||| :::::::::::
365 CCGTCTCCAGCCCATTTCTTGACGGCGGCCACCTCCGCGCACAGTGGCTG 414
94 GluArgSerTyrSerGluHisLeuGlyTyrPheProThrAspLeuPheAl 110
   ||| :::::::::::
415 CTCAAACACATCTATCTCTCTCTTCATTTCTTCACACTCAATATACCAAC 464
110 aCysSerGlu...SerLeuArgAsnGlyAsnGlyLeuGluLeuAsnAla 126
   ::::::::::: ||| :::::::::::
465 TGCACCTGAAGTTCTCCAAAGTCAAGTCAAGCTCAACACTGTAATCTCT 514
126 er...LeuSerGluPheGluLysAsnLysLysIle..... 136
   ||| :::::::::::
515 CCACGTGATCTCTGCTTTCGAAAGGCTGATGATTAACACAGAGAGCCT 564
137 SerLeuLeuHisSerSerLysGluLysLeuArgArgGluArgIleLys 153
   ||| :::::::::::
565 AGTGAACACCATGCTACTGCTGACGACGCGCTAGAGAAAGCTCAACCA 614
153 rCysCysGluGluLeuArgThrLeuLeuProTyrValLysGlyArgLys 170
   ||| :::::::::::
615 ACGGTTATCATTTGAGTCACTGCTTCTTTGTT.....ACGAAA 658
170 snAspAlaIleSerValLeuGluAlaThrValAspTyrValLysTyrIle 186
   ||| :::::::::::
659 TGGATTAAGCCCTCATTTCTTGTCACACCATGAAATATGTCACAGCTTA 708
187 ArgGluLysIleSerPro...AlaValMetAlaGluIleThrGluAlaLe 202
   ||| :::::::::::
709 CGTAAGAAAGTTCAGAGATCTTCAAGCTAGACCAATCAGACGAGGCT.. 756
202 uGlnSerAsnMetArgPheCysLysLysGluGlnThrProIleGluLeu 219
   :::::::::::
757 .....ACGCTGCAGACAA 769
219 erLeuProGlyThrValMetAlaGluArgLysAsnSerValMetSerThr 235
   ::::::::::: ||| :::::::::::
770 AGGATACAGGTAAGGAGGTGTCAGAGGAGG..... 807
236 TyrSerProGluArgGlyLeuGlnPheLeuThrAsnThrCysTrpAsnG 252
   ::::::::::: ||| :::::::::::
808 .....AAGAGAGAAATGAAGATAGAGAGAGAGAGTGTGGTGGAG 848
252 Y.....CysSerThrProAspAlaGluSerSerLeuAspG 264
   ::::::::::: ||| :::::::::::
849 ACAGCAAGATCAGCGCATCTTCGCTCAACGACACATGAA...GAGG 895
264 LuAlaValArgValProSerSerSerAlaSerGluAsnAlaIleGlyAsp 280
   ||| :::::::::::

```

896 AGATAGTCAGACTAGAACTATCATTCAGAGAGTATGATGATGATGAG 945

281ProTyrLys 283

946 CTCAGGTGTCCATACAAA 963

seq_name: /cgn2_6/ptodata/1/ina/5b_COMB.seq:US-08-537-715-3

seq_documentation_block:

; Sequence 3, Application US/08537715

; Patent No. 5910627

; GENERAL INFORMATION:

; APPLICANT: Chuck, George S.

; APPLICANT: Dooner, Hugo K.

; APPLICANT: Courtney-Guterson, Neal

; APPLICANT: Keller, Janis

; APPLICANT: Nijjar, Charanjit S.

; APPLICANT: Ralston, Edward J.

; TITLE OF INVENTION: PH Genes and Their Uses

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/537,715

; FILING DATE: 16-FEB-1996

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/049,282

; FILING DATE: 16-APR-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/US94/04173

; FILING DATE: 15-APR-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Bastian, Kevin L.

; REGISTRATION NUMBER: 34,774

; REFERENCE/DOCKET NUMBER: 012176-003410US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1857 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 481..1632

; OTHER INFORMATION: /note= "ph6 gene cDNA"

; US-08-537-715-3

alignment_scores:

Quality: 108.50

Ratio: 0.714

Percent Similarity: 49.673

Percent Identity: 23.529

alignment_block:

US-09-389-000-2 x US-08-537-715-3

Align seg 1/1 to: US-08-537-715-3 from: 1 to: 1857


```

77 hvValylsThrGluAsnAlaThrGlyProGluGluLeuGlyLeuProLeu 93
   |||||
848 CGGTCCAGCCCATTTCTGAGCGGGGCCACTCGCGCGAGTGCGTG 897
   |||||
94 GluArgSerTyrSerGluHisLeuGlyTyrPheProThrAspLeuLeu 110
   |||||
898 CTCAAAGACATATCTCTGTCATTTCTTCACCTAAATATACCAAC 947
   |||||
110 acysSerGlu...SerLeuArgAsnGlyLeuGluLeuAsnAla 126
   |||||
948 TGCAGCTGAAGTTCTCCAAAGCAGTACGCTACCACGTTGATTCCT 997
   |||||
126 er...LeuSerGluPheGluLysAsnLysLysLeu... 136
   |||||
998 CCACGTGACATCTGCTTCGAAAGGGGTAGTATACACAAAGAGCCCT 1047
   |||||
137 SerLeuLeuHisSerSerLysGlyLysLeuArgGluArgLysTyr 153
   |||||
1048 AGTGAAGAACCATCTCTGCTGAAACGAGCGCTAGAGAAAGCTCAAC 1097
   |||||
153 rcyscysGluGluLeuArgThrLeuLeuProTyrValLysGlyArgLys 170
   |||||
1098 AGCGTTATCATTTTGAAGTCACTTCTCTTTGTT...ACGAAA 1141
   |||||
170 snAspAlaAlaSerValLeuGluAlaThrValAspTyrValLysTyr 186
   |||||
1142 TCGATTAACCTCCATCTTGTGTGACACCATATGATATGCAAGCAGT 1191
   |||||
187 ArgGlyLysIleSerPro...AlaValMetAlaGlnIleThrGluLeu 202
   |||||
1192 CATAAGAAAGTTGAGATCTTGAGAGCTAGACCATCAGACGAGGCT 1239
   |||||
202 ucInSerAsnMetArgPheCysLysGlnGlnThrProIleGluLeu 219
   |||||
1240 .....ACGCTGCACAGAA 1252
   |||||
219 erLeuProGlyThrValMetAlaGlnArgLysSerValMetSerThr 235
   |||||
1253 ACGATACAGTACTGTGAAGGTGTGCAAGAGAGGGT..... 1290
   |||||
236 TyrSerProGluArgGlyLeuGlnPheLeuThrAsnThrCysTrrpAsn 252
   |||||
1291 .....AAGAGGAATGAGAGATGAGAGAGAGTGTGTGAGAG 1331
   |||||
252 Y.....CysSerThrProAspAlaGlnSerSerLeuAsp 264
   |||||
1332 ACAGGCAAGATCAGCGCATCTTCGCCCTCAGACACATGAA...GAG 1378
   |||||
264 LuAlaValArgValProSerSerSerAlaSerGluAsnAlaIleGlyAsp 280
   |||||
1379 ACATAGTCAGAGTAGAATATCATATTCAGAGAGTATGATGATGAGAG 1428
   |||||
281 .....ProTyrLys 283
   |||||
1429 CTCAGGCTGCCATACAA 1446

```

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:us-07-903-710-2

seq_documentation_block:

Sequence 2, Application US/07903710

Patent No. 5302519

GENERAL INFORMATION:

APPLICANT: Eyer,D.E.;Eisenman,R.N.;Blackwood,E.M.;Ayer,D.M.

TITLE OF INVENTION: MAX: A HELIX-LOOP-HELIX ZIPPER PROTEIN THAT FORMS A

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSSEE: Christensen, O'Connor, Johnson and Kindness

STREET: 2800 Pacific First Center, 1420 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98101-2347

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-5.25 inch, 1.2mb storage
COMPUTER: IBM PC/386 Compatible
OPERATING SYSTEM: MS-DOS 4.01
SOFTWARE: Word for Windows-t
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/903,710
FILING DATE: 19920623
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/756,195
FILING DATE: 09/09/91
ATTORNEY/AGENT INFORMATION:
NAME: Sundsmo, John, S.
REGISTRATION NUMBER: 34,446
REFERENCE/DOCKET NUMBER: FHCR-1-6550
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0727 (direct)
TELEFAX: 1-206-224-0779
TELEX: 4938023
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 510 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
DESCRIPTION: Human helix-loop-helix zipper protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens; Eukaryota; Animalia; Metazoa;
ORGANISM: Chordata; Vertebrata; Mammalia; Theria; Eutheria;
ORGANISM: Primates; Haplorhini; Catarrhini; Hominoidea;
IMMEDIATE SOURCE: Human lymphoid B cell Manca cell line,
US-07-903-710-2

```

alignment_scores:

Quality: 108.00 Length: 131
Ratio: 1.301 Gaps: 3
Percent Similarity: 63.359 Percent Identity: 24.427

alignment_block:

US-09-389-000-2 x US-07-903-710-2 ..

Align seg 1/1 to: US-07-903-710-2 from: 1 to: 510

```

140 HisSerSerLysGluLysLeuArgGluArgGlyIleLysTyrCysCysG 156
   |||||
82 CATATGCACTGGAAGAAACGTAGGAGCACATCAAGACAGCTTTCA 131
   |||||
156 ucInLeuArgThrLeuLeuProTyrValLysGlyArgLysAsnAspAla 173
   |||||
132 CAGTTGGGAGACTAGTCCCATCTCCATCCAGAGAGAGAGCATCCGGG 181
   |||||
173 laSerValLeuGluAlaThrValAspTyrValLysTyrIleArgGlyLys 189
   |||||
182 CCCAATCTTAGCAAAAGCCACAGACTATATCCAGATATGCGAAGAAA 231
   |||||
190 lIleSerProAlaValMetAlaGlnIleThrGluAlaLeuGlnSerAsn 206
   |||||
232 .....AACCAACACACACACAGCAAGAT 254
   |||||
206 tarGpHeCysLysGlnGlnThrProIleGluLeuSerLeuProGlyT 223
   |||||
255 TGCAGCTCAAGCGGACAGATGCTCTTGAGAGCAGAGTCCGCGCAC 304
   |||||
223 hvValMetAlaGlnArgLysSerValMetSerThrTyrSerProGlu 239
   |||||
305 TGGAGAAAGCG...AGGTCAAGTGCCTCACTGACACCAATACCCCTCC 351
   |||||
240 ArgGlyLeuGlnPheLeuThrAsn.....ThCysTrr 250
   |||||

```


Tue Jan 23 10:54:45 2001

us-09-389-000-2.rni

Page 17

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: January 19, 2001, 01:45:14 ; Search time 63 seconds

(without alignments)
436.504 Million cell updates/sec

Title: US-09-389-000-2

Perfect score: 2095

Sequence: 1 MYLKVPSSLSAELEAIKL.....MILKAPPKDLISKELAWGF 405

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 6790655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

PTR_66:*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	146	7.0	685	2 T04073	Intensifier 1 prot
2	115.5	5.5	733	1 HHMS86	heat shock protein
3	114	5.4	1788	2 T31095	vitellogenin precu
4	113.5	5.4	732	1 HHMS86	heat shock protein
5	112	5.3	359	2 T47987	hypothetical prote
6	109	5.2	1922	2 T00637	hypothetical prote
7	108.5	5.2	317	2 A48080	basic helix-loop-h
8	108.5	5.2	728	1 HHCH90	heat shock protein
9	108	5.2	151	2 B38431	Myc-binding factor
10	108	5.2	160	2 I50379	gene max protein
11	108	5.2	160	2 S33118	max protein 1 - hu
12	108	5.2	160	2 A38431	Myc-binding factor
13	108	5.2	160	2 A38488	Myn protein - mous
14	108	5.2	160	2 S39792	protein max - rat
15	108	5.2	1210	2 I39410	AF-4 protein, spli
16	108	5.2	1653	2 A36349	clathrin heavy cha
17	107.5	5.1	337	2 S06956	segmentation prote
18	106.5	5.1	1006	2 S76892	hypothetical prote
19	106	5.1	3744	2 S46715	hypothetical prote
20	105.5	5.0	1244	2 S25327	cytoskeleton assem
21	105	5.0	377	2 C64404	hypothetical prote
22	105	5.0	4588	2 T28667	hypothetical prote
23	104.5	5.0	136	2 I51586	dyein beta heavy
24	104.5	5.0	435	2 A46231	Myc binding protei
25	104.5	5.0	486	2 S74319	helix-loop-helix p
26	104.5	5.0	725	2 A41258	RTG3 protein - yea
27	104	4.9	313	2 T15529	a-agglutinin core
28	103	4.9	378	2 A44443	hypothetical prote
29	103	4.9	725	2 UC1468	basic helix-loop-h
30	102.5	4.9	314	2 T06032	hypothetical prote
31	102.5	4.9	1536	1 R8BY53	regulatory protein
32	102	4.9	310	2 T01090	hypothetical prote
33	102	4.9	859	2 T29630	hypothetical prote
34	101.5	4.8	163	2 I51587	heterodimeric part
35	101.5	4.8	329	2 T01333	hypothetical prote
36	100.5	4.8	496	2 S61716	ribose-phosphate p
37	100	4.8	861	2 G02329	replication contro
38	99.5	4.7	3866	2 B48205	All-1 protein -GTE
39	99.5	4.7	3869	2 A48205	All-1 protein +GTE
40	99	4.7	1736	2 T05174	hypothetical prote
41	98.5	4.7	163	2 B56883	transcription regu
42	98.5	4.7	387	2 T40256	hypothetical prote
43	98.5	4.7	793	2 T40285	hypothetical prote
44	98.5	4.7	1091	2 S33850	fibronectin-binding
45	98	4.7	321	2 T15531	hypothetical prote

ALIGNMENTS

RESULT 1
T04073
Intensifier 1 protein - maize
C:Species: Zea mays (maize)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: T04073
R:Burr, F.A.; Burr, B.; Scheffler, B.E.; Blewitt, M.; Wienand, U.; Matz, E.C.
Plant Cell 8, 1249-1259, 1996
A:Title: The maize repressor-like gene intensifier1 shares homology with the r1/b1 mu
A:Reference number: Z15188; MUID:96573194
A:Accession: T04073
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-685 <BUT>
A:Cross-references: EMBL:U57899; NID:g1420923; PIDN:AAB03841.1; PID:g1420924
A:Experimental source: strain W22
C:Genetics:
A:Gene: int1
A:introns: 36/2; 148/2; 180/3; 185/3; 204/3; 343/1; 496/3; 650/3

Query Match 7.0%; Score 146; DB 2; Length 685;

Best Local Similarity 23.5%; Pred. No. 0.0047;
Matches 73; Conservative 46; Mismatches 11; Indels 80; Gaps 14;

QY	89	IGIPQIRSYSEHGYPTDLPACSESLRNGGLELNAS-LSEFEKKKKISLHSSKELR	147
DB	421	LGAPSSSHRSHRG-----EQSSPEPRDDDBGTSRSRGFPVSOTELSHSVLKERR	476
QY	148	REIRKCCQQLRTLLPYVGRKNDASVLEATVDYKIRKISPAVMAQITEALQSNMR	207
DB	477	RETLNNGFAMLSVPEFV-TKMDRASILGDITIEYKQJRRRI-----QELSSRR	525
QY	208	FEKKQQTPIELSLPTVMAQRENSVMSYSPERGQ-----FLTNCWNGCTPDASSL	262
DB	526	LVGSNCK-----TTMAQOPPPAAS-TEERGRQTSGLARAAAGSRAAEASGNS	576
QY	263	DEAVRPPSSASR-----NAIGDPYKTHISSAALSLNLSLTVRKYSKVTPSYDATVNT	315
DB	577	NIGPEPPAAASDTDEVOSTIG-----SDALLELCPIREGILLVMOA-----L	623
QY	316	NGVTSIHLPSAMPVSSFLGTALLQMAR-----RALHI-----PTVCN	354
DB	624	HQRLREITS-----VQASSAGDVLAKLAKAYKEVHGRSSITEVRAIHLIVSSOWICE	679
QY	355	SEGRIRKSTCL 364	
DB	680	-----KNPCL 684	
RESULT	2		
		HHMS86	

heat shock protein 86 - mouse
N:Alternate names: Hsp86
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1990 #sequence, revision 30-Sep-1990 #text_change 22-Jun-1999
C:Accession: B32848; PS0021; A403345
R:Moore, S.K.; Kozak, C.; Robinson, E.A.; Ullrich, S.J.; Appella, E.
J. Biol. Chem. 264, 5343-5351, 1989
A:Title: Murine 86- and 84-kDa heat shock proteins, cDNA sequences, chromosome assignment
A:Reference number: A32848; MUID:89174568
A:Accession: B32848
A:Molecule type: mRNA
A:Residues: 1-733 <MR1>
A:Cross-references: GB:J04633; NID:g194030; PIDN:AA53068.1; PID:g309318
R:Holtmann, T.; Hoveemann, B.
Gene 74, 491-501, 1988
A:Title: Heat-shock proteins, Hsp84 and Hsp86, of mice and men: two related genes encoded
A:Reference number: PS0020; MUID:89232740
A:Accession: PS0021
A:Molecule type: mRNA
A:Residues: 6,'A','8'-242,247-355,'K' <HOE>
A:Cross-references: GB:M3630; NID:g194032; PIDN:AA37868.1; PID:g194033
R:Moore, S.K.; Appella, E.; Villar, C.J.; Kozak, C.A.
Genomics 10, 1019-1029, 1991
A:Title: Mapping of the mouse 86-kDa heat-shock protein expressed gene (hsp86-1) on chro
A:Reference number: A40338; MUID:92009901
A:Accession: A40338
A:Molecule type: DNA
A:Residues: 556-634 <MR2>
A:Cross-references: GB:M57673; NID:g194028; PIDN:AA37867.1; PID:g194029
R:Legagneux, V.; Mergier, V.; Quelard, C.; Barnier, J.V.; Bensaude, O.; Morange, M.
Differentiation 41, 42-48, 1989
A:Title: High constitutive transcription of Hsp86 gene in murine embryonal carcinoma cell
A:Reference number: A37345; MUID:90033873
A:Accession: A37345
A:Molecule type: mRNA
A:Residues: 460-733 <LEG>
A:Cross-references: GB:X16857; NID:g51456; PIDN:CAA34748.1; PID:g51457
C:Comment: In response to temperature stress, to treatment with certain chemicals and am
molecular-weight proteins.
C:Comment: This protein is one of two forms of 80-90 kDa heat shock proteins found in mi
C:Genetics:
A:Gene: HSP86
A:Introns: 586/3
A:Notes: the list of introns may be incomplete
C:Superfamily: heat shock protein 90
C:Keywords: estrogen-induced protein; heat shock; phosphoprotein; steroid receptor comp
F:223-322/Region: highly charged
F:535-570/Region: highly charged

```

Db      613 RIMNAQ-----ALDNDNSMGYMAAKKHLEI-----NPD--HSIIETLR 648
QY      268 VPS-SSASENALGD---PYKTHISSAALSLN--SLHIVRYYSKY-----TPSPIDA 311
Db      649 OKAEADKNDKSRKDYVILLIETALISGFSLEDPQTHANRIYRIKKLGLGIDEDDPYVD 708
QY      312 T--AVTNONISITHEPSAMPV 330
        |   | | | :      | | |
Db      709 TSAAVTE-----MPLL 720

RESULT      3
T31095
vitellogenin precursor - Oreochromis aureus
C:Species: Oreochromis aureus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000
C:Accession: T31095
R:Lim, E.H.; Lam, T.J.; Ding, J.L.
submitted to the EMBL Data Library, August 1997
A:Description: Cloning of full-length Oreochromis aureus vitellogenin cDNA and its de
A:Reference number: Z20978
A:Accession: T31095
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1788 <LIM>
A:Cross-references: EMBL:AF017250; NID:g4102880; PID:g4102881; PIDN:AMD01615.1
C:Superfamily: vitellogenin

```

```
Query Match      5.58; Score 115.5; DB 1; Length 793;  
Best Local Similarity 22.88; Pred.No.0.99;  
Matches 87; Conservative 57; Mismatches 126; Indels 11; Gaps 22;
```

D

```
OY   6 VPSSLSAEELAIKILIRGKK---KNHSLFEVFI--PENFKGCISGHMDIAL---TEP 57  
    |::|||:::||::||::||::||::||::||:  
Db   LPLNTRBMLDQSKLRIKRNLVKKCLEETFLAEDENCKRTKEDEFSNKIKGIHEDS 454  
  
OY   58 LTMEAMSNVVAKIWTTCPSNTKYKTENAIGPELG----PLORSTSEHLGFPT---DLF 109  
    ||::||||  
Db   QNRKLISELRYYT-----SASGDENVSLSDYCITRKHENOXHI-YFITGETRKDY 503  
  
OY   110 ACS---ESLRNGNDLE-----NASLSEPKNKKS-----LLHSSEKRLR 148  
    |||||::|||::|||:  
Db   AMSAFEVLRR-HGLEVIYMIEPIDEIVCQLKEFGEGTLVSATKGELLEPEDEEEKKO 562  
  
OY   149 ERIKYCEQBGLRTLTPYVGGRNDAASVLAEVADVDYKIIREKI-SPAWAQITTELALOSNM 207  
    ::||::||::||::||::||::||:  
Db   EAERTKTFEINLKIM-----KDILEKVEKVAVNSBLVTSPCCIVLTSTVCWTA NME 612  
  
OY   208 FCKNQQTIELSLGTVAQRENSVMSTYSFERGLQFLINTCNMGCSPPDAESSLDNAVR 267  
    |||||::|||::|||:
```

E

```
Db   1613 NIHIIROANGCILNPASHRGLOVFIFGNELRYKVAADMKKGITC-GAGCT--ASNWDGEY 1669  
  
OY   267 RVPSSASENALG-----DPKYTHISSALT-----NSLHTVERYS 303  
    |||||::|||:  
Db   RPSEQVTDIAISTYAHWLSMFCRDSECSIQSVELERYIFEVESKCXYSEPVL 1729  
  
OY   304 KYTPSYDATAVNQNISMTH-LPS 325  
    :|||:  
Db   1730 OCLPGCIPLYRTTVNVGFHCIPS 1752  
  
RESULT 4  
HHHU86  
heat shock protein 90-alpha - human  
MAInterlate names: heat shock protein 86; heat shock protein 89 alpha  
CjSpecies: Homo sapiens (man)  
CjDate: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 22-Jun-1999  
CjAccession: A32319; J00128; PS0020; S06898; J00724; B31420; S67961  
R.Hickey, E.; Brandon, S.E.; Smale, G.; Lloyd, D.; Weber, L.A.  
Mol Cell Biol 9, 2615-2626, 1989  
A>Title: Sequence and regulation of a gene encoding a human 89-kilodalton heat shock  
A:Reference number: A32319; MUID:B9345979
```

Query Match 5.4%; Score 113.5; DB 1; Length 732;
Best Local Similarity 21.8%; Pred. No. 1.4;
Matches 81; Conservative 58; Mismatches 141; Indels 91; Gaps 18;

[illegible]

Db 559 EKKAKFENLCKIN-----KDILEKKREKVVSNRLVTSPPCITVSTYWTANMER 608
 QY 209 CKKOQPIELSLPGTVMAORENSVMSTYSPERGLQFLTNCWNGCSTPDASSLDEAVRY 268
 Db 609 IMKAQ-----ALRDNSTMGYMAKKLEI-----NPD--HSIIETLRQ 644
 QY 269 PS-SSASENAIGD---PYKTHISSAALSIN--SLHTVRYYS--KYTPSYDATAVTNONT 319
 Db 645 KAEADKNDKSVKDLVILVETALLSSGFSLEDPQTHANRITRYIKIKGLGIDEDDTAAEEA 704
 QY 320 SIHLPSAMPV 330
 Db 705 SPAYTEMPPL 715

RESULT 9
 B38431
 Myc-binding factor Max, short form - human
 C:Species: Homo sapiens (man)
 C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 05-Nov-1999
 C:Accession: B38431
 R:Blackwood, E.M.; Eisenman, R.N.
 A:Title: Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding
 A:Reference number: A38431; MUID:91173288
 A:Accession: B38431
 A:Molecule type: mRNA
 A:Residues: 1-151 <BLA>
 A:Cross-references: GB:M64240; NID:q187390; PIDN:AAA6200.1; PID:q187391
 A:Note: the authors translated the codon GAG for residue 94 as Gly

Query Match 5.2%; Score 108; DB 2; Length 151;
 Best Local Similarity 24.4%; Pred. No. 0.37; Mismatches 57; Indels 16; Gaps 3;
 Matches 32; Conservative 26; Mismatches 57; Indels 16; Gaps 3;
 QY 140 HSKKELRERIRKYCCQQLTLLPYVKGKNDASVLEATVDYVKYIREKISPAVNAQIT 199
 Db 19 HNALEKRRDHIDKSPHSLDVSPLQGEKASRAQILDKATEYIQYMRK-----N 69
 QY 200 EALQSMRCKKQQTPIELSLPGTVMAORENSVMSTYSPERGLQFLTN-----TCWNGC 253
 Db 70 HTHQODIDDLKRONALLEQVRALEKA-RSSAQLQTNYPSSDLSLTYNAGSTISAFDGG 128
 QY 254 STPDASSLDE 264
 Db 129 SDSSSESEPEE 139

RESULT 10
 I50379
 gene max protein - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
 C:Accession: I50379
 R:Solleberger, K.G.; Kao, T.L.; Taparowsky, E.J.
 Oncogene 9, 661-664, 1994
 A:Title: Structural analysis of the chicken max gene.
 A:Reference number: I50379; MUID:94119609
 A:Accession: I50379
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-160 <SOL>
 A:Cross-references: GB:U12469; NID:q414723; PIDN:AAA16834.1; PID:q414724
 C:Genetics:
 A:Gene: max
 A:Introns: 12/3; 21/3; 57/3; 99/1

Query Match 5.2%; Score 108; DB 2; Length 160;
 Best Local Similarity 23.7%; Pred. No. 0.4; Mismatches 58; Indels 16; Gaps 3;
 Matches 31; Conservative 26; Mismatches 58; Indels 16; Gaps 3;

QY 140 HSKKELRERIRKYCCQQLTLLPYVKGKNDASVLEATVDYVKYIREKISPAVNAQIT 199
 Db 28 HNALEKRRDHIDKSPHSLDVSPLQGEKASRAQILDKATEYIQYMRK-----N 78
 QY 200 EALQSMRCKKQQTPIELSLPGTVMAORENSVMSTYSPERGLQFLTN-----TCWNGC 253
 Db 79 HTHQODIDDLKRONALLEQVRALEKASSAQLQANV-PAADSSLYTNKSGSTISAFDGG 137
 QY 254 STPDASSLDE 264
 Db 138 SDSSSESEPEE 148

RESULT 11
 S33118
 max protein 1 - human
 C:Species: Homo sapiens (man)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Oct-1999
 C:Accession: S33118
 R:Vaestrik, I.; Koskinen, P.J.; Altalo, R.; Maekelae, T.P.
 Oncogene 8, 503-507, 1993
 A:Title: Alternative mRNA forms and open reading frames of the max gene.
 A:Reference number: S33118; MUID:93149613
 A:Accession: S33118
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-160 <VAE>
 A:Cross-references: EMBL:x66867; NID:g34469; PIDN:CAA47337.1; PID:g34470
 C:Genetics:
 A:Keywords: DNA binding; transcription regulation

Query Match 5.2%; Score 108; DB 2; Length 160;
 Best Local Similarity 24.4%; Pred. No. 0.4; Mismatches 57; Indels 16; Gaps 3;
 Matches 32; Conservative 26; Mismatches 57; Indels 16; Gaps 3;
 QY 140 HSKKELRERIRKYCCQQLTLLPYVKGKNDASVLEATVDYVKYIREKISPAVNAQIT 199
 Db 28 HNALEKRRDHIDKSPHSLDVSPLQGEKASRAQILDKATEYIQYMRK-----N 78
 QY 200 EALQSMRCKKQQTPIELSLPGTVMAORENSVMSTYSPERGLQFLTN-----TCWNGC 253
 Db 79 HTHQODIDDLKRONALLEQVRALEKA-RSSAQLQTNYPSSDLSLTYNAGSTISAFDGG 137
 QY 254 STPDASSLDE 264
 Db 138 SDSSSESEPEE 148

RESULT 12
 A38431
 Myc-binding factor Max, long form - human
 C:Species: Homo sapiens (man)
 C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 23-Mar-1993
 C:Accession: A38431
 R:Blackwood, E.M.; Eisenman, R.N.
 Science 251, 1211-1216, 1991
 A:Title: Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding
 A:Reference number: A38431; MUID:91173288
 A:Accession: A38431
 A:Molecule type: mRNA
 A:Residues: 1-160 <BLA>
 A:Cross-references: GB:M64240
 A:Note: the authors translated the codon GAG for residue 103 as Gly

Query Match 5.2%; Score 108; DB 2; Length 160;
 Best Local Similarity 24.4%; Pred. No. 0.4; Mismatches 57; Indels 16; Gaps 3;
 Matches 32; Conservative 26; Mismatches 57; Indels 16; Gaps 3;
 QY 140 HSKKELRERIRKYCCQQLTLLPYVKGKNDASVLEATVDYVKYIREKISPAVNAQIT 199
 Db 28 HNALEKRRDHIDKSPHSLDVSPLQGEKASRAQILDKATEYIQYMRK-----N 78
 QY 200 EALQSMRCKKQQTPIELSLPGTVMAORENSVMSTYSPERGLQFLTN-----TCWNGC 253
 Db 79 HTHQODIDDLKRONALLEQVRALEKA-RSSAQLQTNYPSSDLSLTYNAGSTISAFDGG 137
 QY 254 STPDASSLDE 264
 Db 138 SDSSSESEPEE 148

Db 28 HNALEKRRDHKIDKSFHSLRDVSPSLQGEKASRAQIILDKATEYIQVMRK-----N 78

QY 200 EALQSNMRFCKKQOTPIELSLPGTVMAQRENSVMSTSPERGLOFLIN-----TCWNGC 253

Db 79 HTHQODIDDLKRONALLQOVRALEKA-RSSAQLQTNPPSSDNLTYTNKAGSTISAFDGC 137

QY 254 STPDASSLSIDE 264

Db 138 SDSSESEPEE 148

RESULT 13

A38488

Myn protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 31-Jul-1991 #sequence_revision 31-Jul-1991 #text_change 16-Feb-1997

C:Accession: A38488

R:Prendergast, G.C.; Lave, D.; Ziff, E.B.

Cell 65, 395-407, 1991

A:Title: Association of Myn, the murine homolog of max, with c-Myc stimulates methylatic

A:Reference number: A38488; MUID:91208676

A:Accession: A38488

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-160 <PRE>

A:Cross-references: GB:M63903

C:Keywords: DNA binding; transcription regulation

Query Match 5.2%; Score 108; DB 2; Length 160;

Best Local Similarity 24.4%; Pred. No. 0.4;

Matches 32; Conservative 26; Mismatches 57; Indels 16; Gaps 3;

QY 140 HSKSEKLRERIKYCCCEQLRTLLPYVKGKNDASVLEATVDYKYIREKISPAVMAQIT 199

Db 28 HNALEKRRDHKIDKSFHSLRDVSPSLQGEKASRAQIILDKATEYIQVMRK-----N 78

QY 200 EALQSNMRFCKKQOTPIELSLPGTVMAQRENSVMSTSPERGLOFLIN-----TCWNGC 253

Db 79 HTHQODIDDLKRONALLQOVRALEKA-RSSAQLQTNPPSSDNLTYTNKAGSTISAFDGC 137

QY 254 STPDASSLSIDE 264

Db 138 SDSSESEPEE 148

RESULT 14

S39792

Protein max - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 10-Nov-1995

C:Accession: S39792

R:Itawa, M.

Biochim. Biophys. Acta 1216, 492-494, 1993

A:Title: Molecular cloning and sequencing of rat Max cDNA: castration-induced expression

A:Reference number: S39792; MUID:94092746

A:Accession: S39792

A:Molecule type: mRNA

A:Residues: 1-160 <12A>

Query Match 5.2%; Score 108; DB 2; Length 160;

Best Local Similarity 24.4%; Pred. No. 0.4;

Matches 32; Conservative 26; Mismatches 57; Indels 16; Gaps 3;

QY 140 HSKSEKLRERIKYCCCEQLRTLLPYVKGKNDASVLEATVDYKYIREKISPAVMAQIT 199

Db 28 HNALEKRRDHKIDKSFHSLRDVSPSLQGEKASRAQIILDKATEYIQVMRK-----N 78

QY 200 EALQSNMRFCKKQOTPIELSLPGTVMAQRENSVMSTSPERGLOFLIN-----TCWNGC 253

Db 79 HTHQODIDDLKRONALLQOVRALEKA-RSSAQLQTNPPSSDNLTYTNKAGSTISAFDGC 137

QY 254 STPDASSLSIDE 264

Db 138 SDSSESEPEE 148

RESULT 15

I39410

AF-4 protein, splice form 3 - human

N:Alternate names: serine/proline-rich FBL protein

C:Species: Homo sapiens (man)

C:Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 29-Aug-1997

C:Accession: I39410

R:Nakamura, T.; Alder, H.; Gu, Y.; Pressad, R.; Canaan, O.; Kamada, N.; Gale, R.P.; L

Proc. Natl. Acad. Sci. U.S.A. 90, 4631-4635, 1993

A:Title: Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute l

A:Reference number: A47440; MUID:93281633

A:Accession: I39410

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1210 <NAK>

A:Cross-references: GB:I13773; NID:g306446; PID:g306447

C:Comment: This protein is one of several to form chimeric fusion proteins with the t

C:Genetics:

A:Gene: GDB:MUT2; AF-4; FBL

A:Cross-references: GDB:136792; OMIM:159557

A:Map position: 4q21-4q21

C:Keywords: alternative splicing

Query Match 5.2%; Score 108; DB 2; Length 1210;

Best Local Similarity 21.4%; Pred. No. 7.4;

Matches 64; Conservative 44; Mismatches 119; Indels 72; Gaps 12;

QY 72 TCPSNTYKTENAT--GEDELGLPIQ-----RSYSEHGLY-----PPTDLFACSES 114

Db 891 TCGDPPKASASTSKSNKDSIPQRVEGKSGRSSEHSGSGDTANPPV-----PS 944

QY 115 LRNGGLNLNALSSEFEKKNKISLH-----SKSEKLRERIKYCCCEQLRTLLPYVK- 166

Db 945 LPNGNS---KPGKQVYFDKQADLHREAKMKQKALMTDRGKAFKLEAVLSIEC 1001

QY 167 -----GRKNDASVLEATVDYKYI-----REKISPAVMAQITEALQSN 205

Db 1002 GIATESSESQSKSAYSVSETVDLIRFIMSLKSPSDATAPQEXIFAVLCMQOSILMA 1061

QY 206 MRFCKKQOTPIELSLPGTVMAQRENSVMSTSPERGLOFLINTCWNGCSTPDASSLSIDEA 265

Db 1062 MFRCKKD--IAIKYSRTLKHFESSSSKVAQAP-----SPCIASTGTGPPSPSPSP 1110

QY 266 VRVSSASASENAIGDPYKTHISSALSLNSLHT--VRYYSKVTPSYD---ATAVTNQN 318

Db 1111 ASSVGSQSSAGSVSSGVAATISIPVTIQMNTSYVITISHVLTAFLPLMDQAEALTKKN 1169

Search completed: January 19, 2001, 10:46:10

Job time: 32456 sec

Tue Jan 23 10:54:45 2001

us-09-389-000-2.rpr

Page 7

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 19, 2001, 03:27:29 ; Search time 62.37 Seconds
(without alignments)
209.702 Million cell updates/sec

Title: US-09-389-000-2
Sequence: 1 MYLKPSSLNLEAIK.....MILKAPKDLISKELATGFG 405

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues
Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	115.5	5.5	732	1 HS9A_MOUSE	P07901 mus musculus
2	113.5	5.4	731	1 HS9A_HUMAN	P07900 homo sapien
3	112.5	5.4	732	1 HS9A_PIG	P02705 sus scrofa
4	110	5.3	160	1 MAX_MOUSE	P28574 mus musculus
5	108.5	5.2	728	1 HS9A_CHICK	P11501 gallus galli
6	108	5.2	160	1 MAX_CHICK	P2162 gallus galli
7	108	5.2	160	1 MAX_HUMAN	P25912 homo sapien
8	108	5.2	160	1 MAX_RAT	P2164 rattus norv
9	108	5.2	1210	1 AF4_HUMAN	P21825 homo sapien
10	108	5.2	1653	1 CLH_YEAST	P21137 saccharomyc
11	107.5	5.1	337	1 HAIR_DROME	P14003 drosophila
12	107.5	5.1	1495	1 MAX1_HUMAN	Q13233 homo sapien
13	107	5.1	378	1 HAIR_DROVI	P29303 drosophila
14	106.5	5.1	732	1 HS9A_CHICK	P29303 drosophila
15	106	5.1	3744	1 YHP9_YEAST	P28811 saccharomyc
16	105.5	5.0	1244	1 SLAI_YEAST	P32790 saccharomyc
17	105	5.0	377	1 Y835_METJA	Q68245 methanococc
18	104.5	5.0	209	1 MAD4_MOUSE	Q60948 mus musculus
19	104.5	5.0	435	1 DPN_DROME	Q26263 drosophila
20	104.5	5.0	486	1 RTG3_YEAST	P28163 saccharomyc
21	104.5	5.0	725	1 AGAL_YEAST	P23233 saccharomyc
22	104	5.0	209	1 MAD4_HUMAN	Q14582 homo sapien
23	103	4.9	725	1 HS9B_CHICK	Q04619 gallus galli
24	102.5	4.9	1536	1 SIN3_YEAST	P22579 saccharomyc
25	101.5	4.8	163	1 MAX_XENLA	O07016 xenopus lae
26	101.5	4.8	723	1 HS9B_MOUSE	P11499 mus musculus
27	100.5	4.8	496	1 KPR5_YEAST	Q12265 saccharomyc
28	100	4.8	861	1 ORC1_HUMAN	Q13415 homo sapien
29	99.5	4.7	3866	1 HRX_MOUSE	P55200 mus musculus
30	99	4.7	1476	1 AT7A_CRIGR	P49045 cricetus
31	98.5	4.7	726	1 HS9A_BRARE	Q90474 brachydanio
32	98	4.7	901	1 MSH5_YEAST	Q12175 saccharomyc
33	97	4.6	1306	1 MSB2_YEAST	P23334 saccharomyc

ALIGNMENTS

RESULT	ID	HS9A_MOUSE	STANDARD	PRT	732 AA.
34	96.5	4.6	446	1 TFE3_MOUSE	O64092 mus musculus
35	96	4.6	416	1 PAX6_COTJA	P47238 coturnix co
36	96	4.6	422	1 PAX6_HUMAN	P26367 homo sapien
37	96	4.6	422	1 PAX6_MOUSE	P32117 mus musculus
38	96	4.6	422	1 PAX6_RAT	P70601 rattus norv
39	95.5	4.6	1121	1 EX5C_HAEIN	P44945 haemophilus
40	95	4.5	363	1 DP3B_AOUAE	O67725 aquifex aeo
41	95	4.5	437	1 PAX6_ORYLA	O73917 oryzae lat
42	95	4.5	666	1 NOD_DROME	P18105 drosophila
43	94.5	4.5	1772	1 MSP1_PLAYO	P13828 plasmodium
44	94	4.5	415	1 PROA_BACSU	P39821 bacillus su
45	94	4.5	575	1 YRAM_HAEIN	P45299 haemophilus

"The human double-stranded DNA-activated protein kinase phosphorylates the 90-kDa heat-shock protein, hsp90 alpha at two NH2-terminal threonine residues." J. Biol. Chem. 264:17275-17280(1989).

- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY (BY SIMILARITY).

- SUBCELLULAR LOCATION: CYTOPLASMIC.

- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@isb-sdb.ch).

CC -----

DR EMBL: J04633; AAA53068.1; -

DR EMBL: M36830; AAA37868.1; ALT_SEQ.

DR EMBL: M57673; AAA37867.1; -

DR EMBL: X16857; CAA34748.1; -

DR PIR: B32848; HHS86.

DR HSSP: P07900; IYES.

DR MGD: MGI:96250; HSP86-1.

DR INTERPRO: IPR001404; -

DR PFAM: PF00183; HSP90; 1.

DR PRINTS: PR00775; HEATSHOCK90.

DR PROSITE: PS00298; HSP90; 1.

DR Chaperone; ATP-binding; Heat shock; Phosphorylation.

KW INIT_MET 0

FT MOD_RES 4 4 PHOSPHORYLATION (BY DS-DNA KINASE).

FT MOD_RES 6 6 PHOSPHORYLATION (BY DS-DNA KINASE).

FT MOD_RES 230 230 PHOSPHORYLATION (BY SIMILARITY).

FT MOD_RES 262 262 PHOSPHORYLATION (BY SIMILARITY).

FT CONFLICT 6 6 T -> A (IN REF. 2).

FT CONFLICT 242 245 MISSING (IN REF. 2).

FT CONFLICT 355 355 R -> K (IN REF. 2).

SEQUENCE 732 AA; 84656 MW; E6345DBEABAC0F1B CRC64;

Query Match 5.5%; Score 115.5; DB 1; Length 732;
 Best Local Similarity 22.8%; Pred. No. 0.46;
 Matches 87; Conservative 57; Mismatches 126; Indels 111; Gaps 22;

QY 6 VPSASAELEAIKIRFEKK--KNTHTLFFYFI--PENFGCISGHMDIAL--TEP 57
 DB 394 LPLNTSREMLQOSKILKVRKNLVKKCLEFTLEADKENVKRYEQFSKNKIKLGIHDS 453
 QY 58 LTMKMSNVKWTCPSTNTVKTENATGPEELGL--PLQSYSEHLGYPT--DLF 109
 DB 454 QNRKRLSELRLYYT-----SASGDEWVSLKDYCTRKNENKHI-YFITGERKQV 502
 QY 110 ACS---ESIRNGNGLEL-----NASISEPEKKNKIS-----LHSSKEKLR 148
 DB 503 ANSAVERLRK-HGLEVIYMIPEIDECYQQLKEFGKTLVSTKGLLELPDEEEKKQ 561
 QY 149 ERIKCCCEGLRLFLPYVKGRNDASVLEATVDYKYYIEKI-SPRVAMQITDALOSNR 207
 DB 562 EEKKTKFENLCKIM-----KDLKKYKVVVSNRLVTSPPCIYTSYGTWANE 611
 QY 208 FCKKOOTPIELSLPGTVMAQRENSVSTSPERGLOFLNTGMCNGSTPDASSLDEAVR 267
 DB 612 RIMKQO-----ALRDNSTMGYMAKKHLEI-----NPD--HSITETLR 647
 QY 268 VPS-SSASENALGD---PYKTHISSAALSLN--SLHTVRYYSKY-----TPSYDA 311
 DB 648 OKAEDAKNDKSVKDLVILLYETALLSSGFSLEDPQTHANRITRMKILGIGIDDPDYVD 707
 QY 312 T--ATNTQNIISHLPSAMPV 330
 DB 708 TSAAYTEE-----MPPL 719

RESULT 2

HS9A_HUMAN STANDARD; PRT; 731 AA.

AC P07900;

DT 01-AUG-1988 (Rel. 08, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE HEAT SHOCK PROTEIN HSP 90-ALPHA (HSP 86).

CN HSP90 OR HSP90A.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

RN [1]

RC SEQUENCE FROM N.A.

RP TISSUE=PERIPHERAL BLOOD LYMPHOCYTES;

RX MEDLINE=89386066; PubMed=2780322;

RA Soeda E., Yokoyama K., Yamazaki M., Akaoji K., Miwa T., Imai T.;

RT "Nucleotide sequence of a full-length cDNA for 90 kDa heat-shock protein from human peripheral blood lymphocytes."

RL Nucleic Acids Res. 17:7108-7108(1989).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=91242090; PubMed=1368637;

RA Yamazaki M., Tashiro H., Yokoyama K., Soeda E.;

RT "Molecular cloning of cDNA encoding a human heat-shock protein whose expression is induced by adenovirus type 12 ELA in HeLa cells."

RL Agric. Biol. Chem. 54:3163-3170(1990).

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE=PLACENTA;

RX MEDLINE=89343979; PubMed=2527334;

RA Hickey E., Brandon S.E., Smale G., Lloyd D., Weber L.A.;

RT "Sequence and regulation of a gene encoding a human 89-kilodalton heat shock protein."

RL Mol. Cell. Biol. 9:2615-2626(1989).

RN [4]

RP SEQUENCE OF 1-311 FROM N.A.

RX MEDLINE=88056312; PubMed=2445630;

RA Hoffmann T., Hoyemann B.;

RT "Cloning and nucleotide sequence of the murine hsp84 cDNA and chromosome assignment of related sequences."

RL Gene 56:29-40(1987).

RN [5]

RP SEQUENCE OF 1-311 FROM N.A.

RX MEDLINE=90076956; PubMed=2591742;

RA Walter T., Drabent B., Krebs H., Tomalak M., Heiss S.;

RT "Cloning and analysis of a human 86-kDa heat-shock-protein-encoding gene."

RL Gene 83:105-115(1989).

RN [6]

RP SEQUENCE OF 1-20. AND PHOSPHORYLATION.

RX MEDLINE=89123325; PubMed=2492519;

RA Lees-Miller S., Anderson C.W.;

RT "Two human 90-kDa heat shock proteins are phosphorylated in vivo at conserved serines that are phosphorylated in vitro by casein kinase II."

RL J. Biol. Chem. 264:2431-2437(1989).

RN [7]

RP PHOSPHORYLATION BY DS-DNA KINASE.

RX MEDLINE=90008887; PubMed=2507541;

RA Lees-Miller S., Anderson C.W.;

RT "The human double-stranded DNA-activated protein kinase phosphorylates the 90-kDa heat-shock protein, hsp90 alpha at two NH2-terminal threonine residues."

RL J. Biol. Chem. 264:17275-17280(1989).

RN [8]

RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 10-222.

RX MEDLINE=97262065; PubMed=9108479;

RA Stebbins C.E., Russo A.A., Schneider C., Rosen N., Hartl F.U., Pavlath N.P.;

RT "Crystal structure of an Hsp90-geldanamycin complex: targeting of a protein chaperone by an antitumor agent."

```

RL Cell 89:239-250(1997).
RN [9]
RX X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 10-222.
RA MEDLINE-99034582; PubMed-9817749;
RA Obermann W.M., Sondermann H., Russo A.A., Pavletich N.P., Hartl F.U.;
RT "In vivo function of Hsp90 is dependent on ATP binding and ATP
RT hydrolysis."
RL J. Cell Biol. 143:901-910(1998).
CC -!- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY
CC -!- (BY SIMILARITY)
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X15183; CA33259.1; -
DR EMBL: X07270; CA330255.1; -
DR EMBL: M27024; AA63194.1; -
DR EMBL: M30626; AA36023.1; -
DR PIR: A32319; HHU086.
DR PIR: B31420; B31420.
DR PIR: J00724; J00724.
DR PDB: 1YER; 22-APR-98.
DR PDB: 1YES; 22-APR-98.
DR PDB: 1YET; 22-APR-98.
DR PDB: 1BYQ; 28-OCT-98.
DR TRANSFAC: T00992; -
DR MIM: 140571; -
DR INTERPRO: IPR001404; -
DR PIR: PF00183; HSP90.1.
DR PRINTS: PR00775; HEATSHOCK90.
DR PROSITE: PS00298; HSP90.1.
KW Chaperone; ATP-binding; Heat shock; Phosphorylation; 3D-structure.
FT INIT_MET 0
FT MOD_RES 4 4 PHOSPHORYLATION (BY DS-DNA KINASE).
FT MOD_RES 6 6 PHOSPHORYLATION (BY DS-DNA KINASE).
FT MOD_RES 230 230 PHOSPHORYLATION.
FT MOD_RES 262 262 PHOSPHORYLATION.
FT CONFLICT 62 62 T -> S (IN REF. 3, 4 AND 5).
FT SEQUENCE 731 AA; 84542 MW; 1249ABCFCE06297C CRC64;
SQ
Query Match 5.4%; Score 113.5; DB 1; Length 731;
Best Local Similarity 21.8%; Pred. No. 0.65;
Matches 81; Conservative 58; Mismatches 141; Indels 91; Gaps 18;
OY 6 VSSLSAELEAIKILRFGK---KNTSLFVFII--PENFGCISGHGMDIAL--TEP 57
DB 393 LPLNISREMLQOSKILKYLKRLVKKCLELTELADKENYKFFQFSKNIKLGIHDS 452
OY 58 LTMKMSNVKWTCPSTVYKTNATGPEELG---PLQSYSEHLGYPTDL----- 108
DB 453 QNRKLSELRLRYT-----SASGDEWVSLKRYCTRMKENQKHYYITGETKDOYA 502
OY 109 -FACSESILRNGLEL-----NASLSEFEKNKKIS-----LLHSSKEKLRE 149
DB 503 NSAFYERLRK-HGLEVIYMIPIDEYCYOQLKEFGKILVSTKRGLELPEDEEEKKOE 561
OY 130 RIKYCEQLRLLPYVKGKKNDAVLEATVDYVYKIREKI-SPAVMAQITELQSNMF 208
DB 562 EKRTFENLCKIM-----KDLLEKVKYVSNRLVSPCCIVSTIGWTANMER 611
OY 209 CKKOOTPIELSLPGTYMAQRENSVMTSPERGLOFLTNCWNGCSTPAESSLDAFV 268
DB 612 IKAQ-----ALRDNSTMGYMAKKHLEI-----NPD--HSIETTLKO 647
OY 269 PS-SSASENAIGD---PYKTHISSAALSLN--SLHTVYYS--KVTPSYDATAVTNOMI 319

```

```

DB 648 KAEADKNKSVKDYLLYETALLSSGFSLEDDPQTHANRYIMKILGIDEDDPTADT 707
OY 320 SIHLPSAMPV 330
DB 708 SAAVTEMPPL 718
RESULT 3
HS9A_PIG STANDARD: PRT; 732 AA.
ID HS9A_PIG
AC 002705;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HEAT SHOCK PROTEIN HSP 90-ALPHA (HSP 86).
GN HS9A OR HSP90A.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN:
RA Huang H.W., Jian S.C., Huang H.C., Lee W.C.;
RT "Complete nucleotide sequence of a porcine cDNA for 90-kDa heat
RT shock protein."
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY
CC -!- (BY SIMILARITY)
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U94395; AAC48718.1; -
DR INTERPRO: IPR001404; -
DR PIR: PF00183; HSP90.1.
DR PRINTS: PR00775; HEATSHOCK90.
DR PROSITE: PS00298; HSP90.1.
KW Chaperone; ATP-binding; Heat shock; Phosphorylation.
FT INIT_MET 0
FT MOD_RES 4 4 PHOSPHORYLATION (BY DS-DNA KINASE)
FT MOD_RES 6 6 PHOSPHORYLATION (BY DS-DNA KINASE)
FT MOD_RES 230 230 PHOSPHORYLATION (BY DS-DNA KINASE)
FT MOD_RES 262 262 PHOSPHORYLATION (BY DS-DNA KINASE)
FT SEQUENCE 732 AA; 84643 MW; 759850062F4E5490 CRC64;
SQ
Query Match 5.4%; Score 112.5; DB 1; Length 732;
Best Local Similarity 21.8%; Pred. No. 0.77;
Matches 81; Conservative 58; Mismatches 141; Indels 91; Gaps 18;
OY 6 VSSLSAELEAIKILRFGK---KNTSLFVFII--BENFGCISGHGMDIAL--TEP 57
DB 394 LPLNISREMLQOSKILKYLKRLVKKCLELTELADKENYKFFQFSKNIKLGIHDS 453
OY 58 LTMKMSNVKWTCPSTVYKTNATGPEELG---PLQSYSEHLGYPTDL----- 108
DB 454 QNRKLSELRLRYT-----SASGDEWVSLDYCTRMKENQKHYYITGETKDOYA 503
OY 109 -FACSESILRNGLEL-----NASLSEFEKNKKIS-----LLHSSKEKLRE 149
DB 504 NSAFYERLRK-HGLEVIYMIPIDEYCYOQLKEFGKILVSTKRGLELPEDEEEKKOE 562
OY 150 RIKYCEQLRLLPYVKGKKNDAVLEATVDYVYKIREKI-SPAVMAQITELQSNMF 208

```

```

DB 563 ERTKTEPNTLCKIM-----KDLLEKVEKVVNSRLVTSQCCVSTGYGTANMER 612
QY 209 CKKQPTIELSLPQTVAAQRENSVYSTSPERGLOFLNTQWNGCSTPDASSIDEAVRY 268
DB 613 IMKAQ-----ALDNRNSTMGYMAKKHLEI-----NPD--HSITTEFLR 648
QY 269 PS-SSASENAIGD---PYTHISSALSLN--SLHVRVRYSS--KVPSPDATVNTQNT 319
DB 649 KAENDKNDKSVKDLVILLYETALLSGFSEDPQTHANRYRMKIKGLGIDEDDPTADS 708
QY 320 SIHUPSAMPV 330
DB 709 SAAVTEMPPL 719

RESULT 4
MAX_MOUSE STANDARD; PRT; 160 AA.
AC P28574;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MAX PROTEIN (MYN PROTEIN) (MYC-BINDING NOVEL HLH/LZ PROTEIN).
GN MAX OR MYN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE-91208676; PubMed-1840505;
RA Prendergast G.C., Lave D., Ziff E.B.;
RT "Association of Myo, the murine homolog of max, with c-Myc stimulates
RT methylation sensitive DNA binding and ras cotransformation.";
RL Cell 65:395-407(1991).
RN (2)
RP STRUCTURE BY NMR OF 70-102 IN COMPLEX WITH MYC.
RX MEDLINE-98347001; PubMed-9680483;
RA Lavigne P., Crump M.P., Gagne S.M., Hodges R.S., Kay C.M., Sykes B.D.;
RT "Insights into the mechanism of heterodimerization from the 1H-NMR
RT solution structure of the c-Myc-Max heterodimeric leucine zipper.";
RL J. Mol. Biol. 281:165-181(1998).
CC -1- FUNCTION: TRANSCRIPTION REGULATOR. FORMS A SEQUENCE-SPECIFIC DNA-
CC BINDING PROTEIN COMPLEX WITH MYC OR MAD WHICH RECOGNIZES THE CORE
CC SEQUENCE 5'-CAC(GA)TG-3'. THE MYC-MAX COMPLEX IS A TRANSCRIPTIONAL
CC ACTIVATOR. WHEREAS THE MAD-MAX COMPLEX IS A REPRESSOR. CPG
CC METHYLATION OF THE RECOGNITION SITE GREATLY INHIBITS DNA BINDING.
CC SUGGESTING THAT DNA METHYLATION MAY REGULATE THE C-MYC/MAX COMPLEX
CC IN VIVO.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. BINDS DNA AS A HETERODIMER WITH MYC OR MAD.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE
CC SPLICING. THE SHORT FORMS DIFFERS BY THE DELETION OF A 9 RESIDUES
CC SEGMENT IN THE N-TERMINAL REGION (BY SIMILARITY).
CC -1- INDUCTION: IN 3T3 FIBROBLASTS BY SERUM.
CC -1- PTM: PHOSPHORYLATED (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M63903; AAA39797.1; -
DR PIR: A38488; A38488
DR PDB: 1A93; 18-NOV-98.
DR RDB: 2A93; 23-FEB-99.
DR TRANSFAC: T00523; -

```

```

DR MGD: MGI:96921; MAX.
DR INTERPRO: IPR001092; -
DR INTERPRO: IPR003015; -
DR PFAM: PFO0010; HLH; 1.
DR PROSITE: PS00038; HELIX-LOOP-HELIX; 1.
KW Nuclear protein; DNA-binding; Transcription regulation;
KW Alternative splicing; Phosphorylation; 3D-structure.
FT DNA_BIND 24 36
FT DOMAIN 39 75
FT DOMAIN 81 102
FT VARSPPLIC 13 21
FT SEQUENCE 160 AA; 18197 MW; CBA5E1894E984F56 CRC64;

Query Match 5.38; Score 110; DB 1; Length 160;
Best Local Similarity 24.4%; Pred. No. 0.14;
Matches 32; Conservative 27; Mismatches 56; Indels 16; Gaps 3;

QY 140 HSKERLRREIRIKYCCQQLRTLLPYKGRKNDASVLEATVDVYKIRKISPAVAQIT 199
DB 28 HNALEKRRDHIDSHSLSDVPSLQGERASQAQLDKATEYIQYRRK-----N 78
QY 200 EALQSNMRECKKOQPTIELSLPQTVAAQRENSVYSTSPERGLOFLTN-----TCWNGC 253
DB 79 DTHQDIDDLKRONALLDQGVRALEKA-RSSAQLQNTYPSDSNLTNKKGTISAFDGG 137
QY 254 STPDASSLDE 264
DB 138 SDDSSESEPE 148

RESULT 5
HSPA_CHICK STANDARD; PRT; 728 AA.
ID HSPA_CHICK
AC P11501;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HEAT SHOCK PROTEIN HSP 90-ALPHA.
GN HSPCA.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC gallus.
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-STOMACH; AND GIZZARD;
RX MEDLINE-89165846; PubMed-2923621;
RA Binart N., Chambraud B., Dumas B., Rowlands D.A., Bigogne C.,
RA Levin J.M., Garnier J., Baulieu E.E., Cateilli M.G.;
RT "The cDNA-derived amino acid sequence of chick heat shock protein Mr
RT 90,000 (HSP 90) reveals a 'DNA like' structure: potential site of
RT interaction with steroid receptors.";
RL Biochem. Biophys. Res. Commun. 159:140-147(1989).
RN (2)
RP SEQUENCE OF 1-85 FROM N.A.
RX MEDLINE-89345085; PubMed-2762125;
RA Vourc' H.C., Binart N., Chambraud B., David J.P., Jerome V.,
RA Balleu E.E., Cateilli M.G.;
RT "Isolation and functional analysis of chicken 90-kDa heat shock
RT protein gene promoter.";
RL Nucleic Acids Res. 17:5259-5272(1989).
CC -1- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```



```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X07265; CA30251.1; -
DR EMBL: X15028; CA33132.1; -
DR PIR: S10880; HCH90.
DR HSSP: P07900; IYES.
DR INTERPRO: IPR001404; -
DR PFAM: PF00183; HSP90; 1.
DR PRINTS: PR00775; HEATSHOCK90.
DR PROSITE: PS00298; HSP90; 1.
KM Chaperone; ATP-binding; Heat shock; Phosphorylation.
FT MOD_RES 230 230 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 250 259 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 728 AA; 84059 MW; 6EC5660FAD5006CE CRC64;

Query Match 5.2%; Score 108.5; DB 1; Length 728;
Best Local Similarity 21.6%; Pred. No. 1.5;
Matches 80; Conservative 58; Mismatches 142; Indels 91; Gaps 18;

QY 6 VPSSASAELEAKILIRGKK---KNTHSLFVFIT--PENFGCISGGMIAL---TEP 57
DB 390 LPLNTSRMLQOSKTLKYIRKMLVKKCLEFTLEADENKNTKFTQSKNIKLIHEDS 449
QY 58 LTMKMSVVKYWTCPSTVNTKTNATGPEELG---PLQSYSEHLGYPTDL----- 108
DB 450 QNRKLTSLRLRYT-----SASGDEWSLMDYCRMKENKHYIITGEEKDOYA 499
QY 109 -FAGESLIRNGDEL-----NALSEFEKNKTKIS-----LHSSKEKLRE 149
DB 500 NSAFERLRK-HGLEVIYWIPEIDECYVOQLKEFEGKTLVSTKGLLEPEDEEKKKQE 558
QY 150 RIKYCCEDRLTPYVGRKNDASVLEATVDYVYIREKI-SPAVMAQITFALQSNMF 208
DB 559 EKKAFENLCKTM-----KYLEKKYKVVVSNRLVTSPPCITYTSTGTANNER 608
QY 209 CKKOQPIELSLPTVMAQRENSVMSTYSPERGLOFLNTCWNGSTPDASSLDAVAV 268
DB 609 IMKAQ-----ALRNSTMGYMAKKHLEI-----NPD--HSIETLRLQ 644
QY 269 PS-SSASENAIGD---PKTHISSAALSIN--SLHTVAYIS--KYTPSYDTANTMONI 319
DB 645 KAEADKNKSVKVDLYLLYETALLSSGSELEDPQTHANRYIMKILGIGIDEDTAABA 704
QY 320 SIHLPSAMPV 330
DB 705 SPATVEEMPPL 715

RESULT 6
MAX_CHICK STANDARD: PRT: 160 AA.
AC P52162;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MAX PROTEIN.
GN MAX.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauromorpha; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-FIBROBLAST.
RX MEDLINE=94119609; PubMed=8290277;
RA Solenderberger K.G., Kao T.L., Taparowsky E.J.;
RT "Structural analysis of the chicken max gene."
CC Oncogene 9:661-664(1994).
CC -!- FUNCTION: TRANSCRIPTION REGULATOR. FORMS A SEQUENCE-SPECIFIC
CC DNA-BINDING PROTEIN COMPLEX WITH MYC OR MAD WHICH RECOGNIZES
CC THE CORE SEQUENCE 5'-CAC[CA]TG-3'. THE MYC-MAX COMPLEX IS A

```

```

CC TRANSSCRIPTIONAL ACTIVATOR, WHEREAS THE MAD-MAX COMPLEX IS A
CC REPRESSOR (BY SIMILARITY).
CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. BINDS DNA AS A HETERODIMER WITH MYC OR MAD.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE
CC SPLICING. THE SHORT FORM DIFFERS BY THE DELETION OF A 9 RESIDUES
CC SEGMENT IN THE N-TERMINAL REGION.
CC -!- PTM: PHOSPHORYLATED (PROBABLY).
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L12469; AAA16834.1; -
DR HSSP: P25912; 1AN2.
DR INTERPRO: IPR001092; -
DR INTERPRO: IPR003015; -
DR PFAM: PF00010; HLH; 1.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
KM Nuclear protein; DNA-binding; Transcription regulation;
KW Alternative splicing; Phosphorylation.
FT DNA_BIND 24 36 BASIC DOMAIN.
FT DOMAIN 39 75 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 81 102 LEUCINE-ZIPPER (BY SIMILARITY).
FT VARSPLOC 13 21 MISSING (IN SHORT ISOFORM) (BY
FT SIMILARITY).
SQ SEQUENCE 160 AA; 18184 MW; F09B16D1AB379A7E CRC64;

Query Match 5.2%; Score 108; DB 1; Length 160;
Best Local Similarity 23.7%; Pred. No. 0.2;
Matches 31; Conservative 26; Mismatches 58; Indels 16; Gaps 3;

QY 140 HSSKEKLREKRYCCEDRLTPYVGRKNDASVLEATVDYVYIREKISPAMQIT 199
DB 28 HNALEKRRDRHKDSFSLRDSVPSLGEKASRAQIDLKATEIYIDYMRK-----N 78
QY 200 EALQSNMRFCKKQPIELSLPTVMAQRENSVMSTYSPERGLOFLN-----TCWNGC 253
DB 79 HTHQODIDDLKRNALLDQVRALEKARSSAQQLANY-PADSSLYTPKGSSTISAFDGG 137
QY 254 STPDASSLDE 264
DB 138 SDSSSDSEPE 148

RESULT 7
MAX_HUMAN STANDARD: PRT: 160 AA.
AC P25912; P52163;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MAX PROTEIN.
GN MAX.
OS Homo sapiens (Human), and Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-HUMAN.
RX MEDLINE=91173288; PubMed=2006410;
RA Blackwood E.M., Eisenman R.N.;
RT "Max: a helix-loop-helix zipper protein that forms a
RT sequence-specific DNA-binding complex with Myc."
RT Science 251:1211-1217(1991).

```


CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: D14447; BAA03337.1; -
 DR EMBL: D14448; BAA03338.1; -
 DR HSSP: P25912; IAN2
 DR INTERPRO: IPR001092; -
 DR INTERPRO: IPR003013; -
 DR PFM: PF00010; HLH; 1.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
 KW Nuclear protein; DNA-binding; transcription regulation;
 KW alternative splicing; phosphorylation.
 FT DNA_BIND 24 36 BASIC DOMAIN.
 FT DOMAIN 39 75 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
 FT DOMAIN 81 102 LEUCINE-ZIPPER (BY SIMILARITY).
 FT VARSPIC 13 21 MISSING (IN SHORT ISOFORM).
 SQ SEQUENCE 160 AA; 18272 MW; 86CB1A137727A57A CRC64;

Query Match 5.2%; Score 108; DB 1; Length 160;
 Best Local Similarity 24.4%; Pred. No. 0.2;
 Matches 32; Conservative 26; Mismatches 57; Indels 16; Gaps 3;

QY 140 HSSKRLERIKKCCCEQLRTLLPYVKGRKNDASVLEATVDYVYKREKISPAVMAQIT 199
 Db 28 HNALERRRRDRIKQNALTEQVRALEKA-RSSAQLOTNTPSSDNSLYTNAKGITISAFDGG 78

QY 200 EALOSNRFECKKQJOTPIELSLPGTVMAQRENSVMSTYSPERGLOFLTN-----TCWNGC 253
 Db 79 HTHQODDIDLKQNALTEQVRALEKA-RSSAQLOTNTPSSDNSLYTNAKGITISAFDGG 137

QY 254 STPDSESLIDE 264
 Db 138 SSSSESESEPEE 148

RESULT 9
 AF4_HUMAN STANDARD; PRT; 1210 AA.

AC P51825;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE AF-4 PROTEIN (FEL. PROTEIN).
 GN MLT2 OR AF4 OR FEL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE:93281633; PubMed-8506309;
 RA Nakamura T., Alder H., Gu Y., Prasad R., Canaan O., Kamada N.,
 RA Gale R.P., Lange B., Crist W.M., Nowell P.C., Croce C.M.,
 RA Canaan E.;
 RA "Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in
 RT acute leukemia share sequence homology and/or common motifs";
 RT Proc. Natl. Acad. Sci. U.S.A. 90:4631-4635(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93184301; PubMed-8443374;
 RA Morrissey J., Tkachuk D.C., Milatovich A., Francke U., Link M.,
 RA Cleary M.L.;
 RA "A serine/proline-rich protein is fused to HRX in t(4;11) acute
 RT leukemias";
 RT Blood 81:1124-1131(1993).
 RL -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- DISEASE: INVOLVED IN A T(4;11)(Q21;Q23) CHROMOSOMAL TRANSLOCATION
 CC IN ACUTE LEUKEMIAS CAUSING FUSION TO THE MLT/HRX GENE PRODUCT
 CC WHICH CONTAINS DNA-BINDING MOTIFS RESULTING IN A ROGUE ACTIVATOR

CC PROTEIN.
 CC -1- SIMILARITY: TO LAF4 AND FMR2 (OX19).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: L13773; AAA58360.1; -
 DR EMBL: L25050; AAA36642.1; -
 DR MIM: 159557; -
 KW Nuclear protein; Chromosomal translocation; proto-oncogene.
 FT DOMAIN 483 492 POLY-SER.
 FT DOMAIN 835 843 POLY-SER.
 FT DOMAIN 866 869 POLY-PRO.
 FT DOMAIN 871 874 POLY-SER.
 FT CONFLICT 46 46 K -> R (IN REF. 2).
 FT CONFLICT 624 624 E -> G (IN REF. 2).
 FT CONFLICT 899 905 SASSTKS -> VPAVPRV (IN REF. 2).
 FT CONFLICT 928 929 EH -> AD (IN REF. 2).
 FT CONFLICT 999 999 I -> N (IN REF. 2).
 FT CONFLICT 1096 1096 A -> AR (IN REF. 2).
 FT CONFLICT 1140 1140 N -> I (IN REF. 2).
 STNCVTLANSLVDVHYRGRGFOLOELTKTP -> RQ
 MCAVPESTAVMWTCTHHDVSESYKN (IN REF. 2).
 SQ SEQUENCE 1210 AA; 131421 MW; F0E334DF8C2EF04 CRC64;

Query Match 5.2%; Score 108; DB 1; Length 1210;
 Best Local Similarity 21.4%; Pred. No. 3.3;
 Matches 64; Conservative 44; Mismatches 119; Indels 72; Gaps 12;

QY 72 TCEPNVKTENAT--GPEELGAPQ-----RSYSEHLGY-----PPTDLFACSES 114
 Db 891 TCGDDPKRSASSTKSNHSDSIIPKORVEGKGRSSSHKSGDGTANFPV-----PS 944

QY 115 LRNGGLELNSLSEPEENKKSILH-----SKKELRERIKKCCCEQLRTLLPYVK 166
 Db 945 LPNGNS--KPKQPVAFKQKQADLHMRKAKKQKQKALMDRGAFAKYLEAVLSFEC 1001

QY 167 -----GRKNDASVLEATVDYVYK-----REKISPAVMAQITREALQSN 205
 Db 1002 GIARESSQSKSAVSYSEVVDLKFIMLSKFSDDATPQKIFAVLCRCQSLNMA 1061

QY 206 MRECKKQOTPIELSLPGTVMAQRENSVMSTYSPERGLOFLTNCGWNGCSTPDSESLIDEA 265
 Db 1062 MFRCKKD--IAIKYSRTLNHFSSSKVAQAP-----SPCIASTGTFPSPLSPMPSP 1110

QY 266 VRVSSASASEMAIDDPKTHSSAALSLSLHT--VRIYSKYVTSYD--ATAVTNQN 318
 Db 1111 ASSVSQSSAGVSGVAVATISPTVIQNTSSYVITSHVLAFLDMEQAEALTRKN 1169

RESULT 10
 CLH_YEAST STANDARD; PRT; 1653 AA.

AC P22137;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE CLATHRIN HEAVY CHAIN.
 GN CHC1 OR YGL206C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAT-8288C;
 RX MEDLINE-91093343; PubMed-1898742;
 RA Lemmon S.K., Pellicena-Palle A., Conley K., Freund C.L.;

[illegible]

DR EMBL: X15904; CAA34018.1; -
 DR EMBL: X15905; CAA34019.1; -
 DR PIR: S06956; S06956.
 DR TRANSFAC: T00345; -
 DR FLYBASE: FBgn0001168; h.
 DR INTERPRO: IPR001092; -
 DR INTERPRO: IPR003015; -
 DR PFAM: PF00010; HLH.1.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX.1.
 DR Nuclear protein; Developmental protein; Pair-rule protein;
 DR DNA-binding; Transcription regulation; Repressor.
 FT DNA_BIND 32 44
 FT DOMAIN 45 89
 FT DOMAIN 149 157
 FT DOMAIN 222 237
 FT DOMAIN 241 250
 FT VARIANT 292 292
 FT DOMAIN 334 337
 SO SEQUENCE 337 AA; 36995 MW; 6D2ECATFF2D56C0B CRC64;
 WRPW MOTIF (REQUIRED FOR ACTIVITY).
 S -> P.
 P -> P.
 Query Match 5.1%; Score 107.5; DB 1; Length 337;
 Best Local Similarity 23.7%; Pred. No. 0.62; Mismatches 84; Indels 55; Gaps 10;
 Matches 53; Conservative 32; Mismatches 84; Indels 55; Gaps 10;
 QY 128 SEFEKNNKISLHSSKEKLRERIKYCEQLRTL-----PYVGRKNDASVLEATVD 181
 DB 30 SDRSRNKPI-----MEKRRRRRINNCINELKTLILDATKDPARHSKLEKADILEKTVK 83
 QY 182 YVYIYREK---ISPAVMAQTLEALQSNMRCCKQQTPIELSLPGTVMAQENSVMTSYSP 238
 DB 84 HLOELDROQAAMQADPKIVNKFAGFADCVNEVS---RFPGLIEPAQR-----129
 QY 239 ERGLQFLNTCMGCGST-----PDASSLEAVVPSSSA---SENAGDPYKTHIS 287
 DB 130 RRLQHLHN-CINGVKTELHQOORQOOSIHQMPLSPSSPEODSQCAAPYLFGLQ 168
 QY 288 SAALSL--NSLHTVRYYSKVPSTYDATVNTONISIHLSAMP 328
 DB 189 QFASGYFLPMGMQVI-----PTKLPGNSIALVLPQSLP 221
 RESULT 12
 M3L_HUMAN STANDARD; PRT; 1495 AA.
 AC 013233;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE MITOGEN-ACTIVATED PROTEIN KINASE 1 (EC 2.7.1.-)
 DE (MAPK/ERK KINASE KINASE 1) (MEK KINASE 1) (FRAGMENT).
 GN MAP3K1 OR MAPKKK1 OR MEKK1 OR MEKK.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN (1)
 RP SEQUENCE FROM N.A.
 RA MEDLINE-9902611; PubMed-9808624;
 RA Xia Y., Wu Z., Su B., Murray B., Karin M.;
 RT "JNK1 organizes a MAP kinase module through specific and sequential
 RT interactions with upstream and downstream components mediated by its
 RT amino-terminal extension.";
 RL Genes Dev. 12:3369-3381(1998).
 RN (2)
 RP SEQUENCE OF 1221-1257 FROM N.A.
 RC TISSUE-LEUCOCYTE;
 RX MEDLINE-96169565; PubMed-8597633;
 RA Vink B.S., Kay E.S., Fiedorek F.T.;
 RT "Mapping of the MEK Kinase gene (MeKk) to mouse chromosome 13 and
 RT human chromosome 5.";
 RL Mamm. Genome 6:782-783(1995).
 CC -1- FUNCTION: CAN PHOSPHORYLATE AND ACTIVATE MAPK 1 AND MAPK 2
 CC (MEK1/MEK2) WHICH LEADS TO PHOSPHORYLATION OF MAP KINASES. IT IS

CC ALSO A HIGHLY EFFICIENT ACTIVATOR OF THE JNK CASCADE.
 CC -1- PTM: AUTOPHOSPHORYLATED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC MAP KINASE KINASE KINASE SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 2 PH DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.slb.ch).
 CC -----
 CC EMBL: AF042838; AAC97073.1; -
 CC EMBL: U29671; AAB05828.1; -
 CC DR SWISS-2DPAGE; Q13233; HUMAN.
 DR MIM: 600982; -
 DR INTERPRO: IPR000719; -
 DR INTERPRO: IPR002290; -
 DR PFAM: PF00069; PKINASE.1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP.1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST.1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM.1.
 DR Transferrase; Serine/threonine-protein kinase; ATP-binding;
 KW Phosphorylation.
 FT NON_TER 1 1
 FT DOMAIN 6 10
 FT DOMAIN 17 22
 FT DOMAIN 403 412
 FT DOMAIN 430 442
 FT DOMAIN 430 442
 FT DOMAIN 633 740
 FT DOMAIN 925 932
 FT DOMAIN 1165 1170
 FT DOMAIN 1199 1202
 FT DOMAIN 1226 1481
 FT DOMAIN 1232 1239
 FT NE_BIND 1255 1255
 FT BINDING 1352 1352
 FT ACT_SITE 1352 1352
 SO SEQUENCE 1495 AA; 163653 MW; 2D6ECEB52E7AFCAB CRC64;
 Query Match 5.1%; Score 107.5; DB 1; Length 1495;
 Best Local Similarity 24.1%; Pred. No. 4.9; Mismatches 95; Indels 65; Gaps 10;
 Matches 62; Conservative 35; Mismatches 95; Indels 65; Gaps 10;
 QY 129 EFEKNNKISL-----HSSKEKLRERIKYCEQLRTLPPYVGRKNDASVLEA 178
 DB 769 EIRYKLLSLTLTALDSIDNSHMGRLSR-RYLSARAVTVVPAVFSKLEMLSVSY 827
 QY 179 TVDYVYIYREKISPAVMAQTLEALQSNMRCCKQQTPIELSLPGTVMAQENSV----- 232
 DB 828 STHTFRRRRLMAYADEVEIAEAIQ-----LGVEDTLQROOHNSFCRHLFP 873
 QY 233 MSTYSPERGLQF-----LTYTCNCGSTPDASSLDEAVY-----PSSASAEAL- 278
 DB 874 TTIWKPRVPLECTVLEKGTGKLCATKLSASSSEISERLARIISVGPSSSTTTTTE 933
 QY 279 -----GDPYKTHISSAALSLNS--LHTVRYYSKYTPSYDATVNTON-----I 319
 DB 934 QPKPMWOTKGRPHSGLNSPLSHSOLMFPALSTSSSTPSVPAGTANDVSNRLOGFI 993
 QY 320 SIHLPSAMPVVS-SFSL 335
 DB 994 PCRIPIASAPQTKRFSL 1010
 RESULT 13
 HAIR_DPROVI STANDARD; PRT; 378 AA.
 ID HAIR_DPROVI
 AC P29303;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)


```

Db      612 RIRAKQ-----ALRDNSTMGYMAARKHLEI-----NPD--HSLIETLR 647
QY      268 VPS-SSASENAIGD---PYKTHISSAALSLN--SLHTVRYSKY-----TPSDA 311
Db      648 OKAEDKNDKSVKDLVILLYETALISSGFSLEDPOTFANRIRYKIKGLGIDEDDPYDD 707
QY      312 T--AVTNONISIHLPSPMPV 330
Db      708 TSAVTEE-----MPL 719

```

```

RESULT 15
YHP9_YEAST
ID YHP9_YEAST STANDARD: PRT: 3744 AA.
AC P38811:

```

```

DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HYPOTHEETICAL 433.2 KDA PROTEIN IN HXT5-NRK1 INTERGENIC REGION.
GN YHR099W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RX MEDLINE-94378003; PubMed-8091229;
RA Johnston M., Andrews S., Brickman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favello A., Fulton L., Gattung S., Giesel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Treaskis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII."
RL Science 265:2077-2082(1994).
CC -1- SIMILARITY: STRONG, TO S.POMBE SPAC1F5.11C.
CC -1- SIMILARITY: BELONGS TO THE P13/P14-KINASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC
CC EMBL; U00060; AAB68923.1; -
CC
CC PIR; S46715; S46715.
CC
CC SGD; S0001141; YHR099W.
CC
CC Hypothetical protein; Transferase; Kinase.
CC
CC DOMAIN 3414 3744 P13K/P14K.
CC
CC SEQUENCE 3744 AA; 433171 MW; AE358676F5D5777 CRC64;

```

```

Query Match 5.1%; Score 106; DB 1; Length 3744;
Best Local Similarity 23.0%; Pred. No. 22;
Matches 58; Conservative 40; Mismatches 114; Indels 40; Gaps 9;

```

```

QY 123 LNASLSEFEKNNKISLHSSKEKLRERIKYCC-----EQLRTLLPYVGRKNDAA 173
Db 1821 LKSYLVEDKCKRKLHNLKLNKSNALAVDLHDHDLFRELQLSAIFIKA---DPE 1877
QY 174 SVLEATVDYVYIRE--KISPAVMAQITEALQSNMRCKKOOTPIELSLPOTVNAQRENS 231
Db 1878 IIAEIKKDIKFCWNFKILEDTLILKQSAVLYTS--YFISKDFPIKVTQYVVALRRSSH 1935
QY 232 VMSYSPERGLOFLT---NTCMNGCSTPDASSLDEAVRVPSSASASENAIGDPTKTHISS 288
Db 1936 VEARYLVKQSLDVLTPVYHERMNAAGTPTWIMVYKRYVENSSQNNIL---YQFLISH 1992

```

```

QY 289 AALSLNS-----LHTVRYSKVTPSDATAVTNONISIHLPSPMPVSSSLGTALLGWA 343
Db 1993 PDLFPNSRDLFISNIHNMKI-----TFMSNSNSDSH-----TLAIDLASLIYWE 2039
QY 344 RRAHPIPYVCS 355
Db 2040 NKTLEITVNMNT 2051

```

```

Search completed: January 19, 2001, 10:52:08
Job time: 26679 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 19, 2001, 02:57:44 ; Search time 65.98 Seconds
(without alignments)
719.448 Million cell updates/sec

Title: US-09-389-000-2
Sequence: 1 MWLKYPSLSAELEAIKIL.....MIKAPPKDLISKELAEWGF 405

Scoring table: BLOSUM62
Gapop 10.50, Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_15:*
1: sp_archaea*
2: sp_bacteria*
3: sp_fungi*
4: sp_human*
5: sp_invertebrate*
6: sp_mammal*
7: sp_mhc*
8: sp_organelle*
9: sp_phage*
10: sp_plant*
11: sp_rodent*
12: sp_virus*
13: sp_vertebrate*
14: sp_unclassified*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1699	81.1	425	Q9NX45	Q9NX45 homo sapien
2	146	7.0	685	Q41875	Q41875 zea mays (m
3	121	5.8	367	Q65552	Q65552 arabidopsis
4	119	5.7	452	Q49771	Q49771 homo sapien
5	119	5.4	452	Q9U149	Q9U149 homo sapien
6	114	5.4	1788	Q9Y6K0	Q9Y6K0 oreochromis
7	113.5	5.4	539	Q75322	Q75322 homo sapien
8	113	5.4	311	Q9S0S3	Q9S0S3 arabidopsis
9	112	5.3	350	Q9ZUG9	Q9ZUG9 arabidopsis
10	112	5.3	359	Q9M270	Q9M270 arabidopsis
11	111	5.3	254	Q9XEF0	Q9XEF0 arabidopsis
12	110	5.3	662	Q9ZPY8	Q9ZPY8 arabidopsis
13	109	5.2	339	Q9QUS4	Q9QUS4 mus musculu
14	109	5.2	1708	Q94895	Q94895 homo sapien
15	109	5.2	1773	Q9Y6B8	Q9Y6B8 homo sapien
16	109	5.2	1922	Q43355	Q43355 homo sapien
17	109	5.2	3595	Q9UCH3	Q9UCH3 homo sapien
18	109	5.2	3908	Q9UC04	Q9UC04 homo sapien
19	109	5.2	3911	Q99996	Q99996 homo sapien

20	107.5	5.1	337	Q9VSN8	Q9VSN8 drosophila
21	107	5.1	415	Q23192	Q23192 arabidopsis
22	106.5	5.1	1006	P74686	P74686 synechocyst
23	105.5	5.0	297	Q9LSQ3	Q9LSQ3 arabidopsis
24	105.5	5.0	446	Q9NVH3	Q9NVH3 homo sapien
25	105.5	5.0	796	Q9P217	Q9P217 homo sapien
26	105	5.0	460	P91751	P91751 loligo opal
27	105	5.0	4588	Q94709	Q94709 parametium
28	104.5	5.0	204	Q9W6H1	Q9W6H1 xenopus lae
29	104.5	5.0	435	Q9V384	Q9V384 drosophila
30	104	5.0	313	Q18054	Q18054 caenorhabdi
31	103	4.9	297	Q9SA66	Q9SA66 arabidopsis
32	103	4.9	423	Q9SKX6	Q9SKX6 arabidopsis
33	103	4.9	2627	Q99973	Q99973 homo sapien
34	103	4.9	3899	Q9Y6X2	Q9Y6X2 homo sapien
35	102.5	4.9	314	Q9U072	Q9U072 arabidopsis
36	102.5	4.9	524	Q9SBC5	Q9SBC5 arabidopsis
37	102.5	4.9	848	Q9RAT4	Q9RAT4 lactococcus
38	102.5	4.9	1492	P70705	P70705 rattus norv
39	102	4.9	310	Q22768	Q22768 arabidopsis
40	102	4.9	337	Q9UBP5	Q9UBP5 h basic-hel
41	102	4.9	859	Q21401	Q21401 caenorhabdi
42	101.5	4.8	329	Q81306	Q81306 arabidopsis
43	101.5	4.8	416	Q78234	Q78234 chrysosplen
44	101.5	4.8	416	Q78237	Q78237 chrysosplen
45	101.5	4.8	979	Q9VEN2	Q9VEN2 drosophila

ALIGNMENTS

RESULT 1
Q9NX45 PRELIMINARY; PRT; 425 AA.
ID Q9NX45
AC Q9NX45
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TEMBLrel. 15, Last annotation update)
DE CNA FLJ20449 FIS, CLONE KAT05575.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Watanabe K., Kunagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isegaki T., Sugano S.;
RT "NEDO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK000456; BAA91175.1; -
SQ SEQUENCE 425 AA; 46922 MW; 777E57231F0D6BDD CRC64;

Query Match	81.1%	Score 1699;	DB 4;	Length 425;
Best Local Similarity	89.4%	Pred. No. 4.1e-126;		
Matches 338;	Conservative 5;	Mismatches 11;	Indels 24;	Gaps 4;
QY 1	MWLKVPSSLSAELEAIKILRFQKKNTSLFVFIIPENFKGICSHGMDIALTEPLTM 60			
DB 67	MWLKVPSSLSAELEAIKILRFQKKNTSLFVFIIPENFKGICSHGMDIALTEPLTM 126			
QY 61	EKMSNVKWTTCSPNVKTEENATGPELGLPLDRSYSEHLGYPTDLFACSESLRANG 120			
DB 127	EKMSNVKWTTCSPNVKTEENATGPELGLPLDRSYSEHLGYPTDLFACSESLRANG 186			
QY 121	LELNASISEFEKKKIKSLHSSKKELRERIKYCCEDLRILLPVYKRRKNDASVLEATV 180			
DB 187	LELNASISEFEKKKIKSLHSSKKELRERIKYCCEDLRILLPVYKRRKNDASVLEATV 246			
QY 181	DYVYTIKESIPAYMAOITELALOSNMRFCKKQOPIELSLPGYMAORENSVMTYSPE 240			
DB 247	DYVYTIKESIPAYMAOITELALOSNMRFCKKQOPIELSLPGYMAORENSVMTYSPE 306			

```

QY 241 GLOFLNTGNCSTPDASSLDEAVRPSSASANAIDGPYKTHISSAALSTNSHTYR 300
DB 307 GLOFLANTGNCSTPDASSLDEAVRPSSASANAIDGPYKTHISSAALSTNSHTYR 366
QY 301 YSKVTPSYDATAVTNQNISIHLPVSSPSTLGLGMMARALHPTVCNS-FGRI 359
DB 367 YSKVTPSYDATAVTNQNISIHLPVSSPSTLGLGMMARALHPTVCNS-FGRI 411
QY 360 KST---CLKFTLSTIYWA 374
DB 412 CTHPNCLO-----QFWA 424

RESULT 2
QY 041875 PRELIMINARY: PRT: 685 AA.
AC 041875;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE REPRESSOR-LIKE PROTEIN (INI).
GN INI.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
OX NCBI_TaxId=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W22;
RA Burr F.A.; Burr B.; Scheffler B.E.; Biewitt M.; Wienand U.; Matz E.C.;
RL Plant Cell 0.0-0(0).
DR EMBL: U57899; AAB03841.1;
DR MENDEL: 11294; Zeama.1300; 11294.
DR INTERPRO: IPR001092; -.
DR PFM: PF00010; HLH; 1.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN.1.
SQ SEQUENCE 685 AA; 74698 MW; FDC67F8E95F21F79 CRC64;

```

```

Query Match 7.0%; Score 146; DB 10; Length 685;
Best Local Similarity 23.5%; Pred. No. 0.0027;
Matches 73; Conservative 46; Mismatches 111; Indels 80; Gaps 14;

QY 89 LGPLPORSSEHIGYPTDIFACSESLRNGCIELNAS-LSPEKKKKSILHSSSEKIR 147
DB 421 LGAPSSSSSHSRHG---EVQSSSPRRDDEGTSRSRGPVPSQTELSASHVLEKERR 476
QY 148 RRIKRCCEQLRLLPYVGRKNDASVLEATVDYKIRKISPAVMAQITFALQSNR 207
DB 477 RKLNGEFLMLSLVPEV--TKMDRASILGDIIEYKOLRRK-----QELSSRR 525
QY 208 FCKKOQTPIELSLPTVMAORENSVMSTYSPERGLO-----FLTNTGNCSTPDASSL 262
DB 526 LVGSNKG-----TTMAQOPPPPAAS--TEERGRQTSGGYLARAAGSRAEASGNS 576
QY 263 DEAVRPSSASE-----NAIGDPYKTHISSAALSLNLTVRYSVYTSYDATAVT 315
DB 577 NGEEPPAAASTDTEVOVSIIG-----SPALLELCRPHREGILLHYMQA-----L 623
QY 316 NONISIHLPASMPVSSFLGALLGMAR-----RAHHI-----PTVCN 354
DB 624 HOELRELEITS---VOASSAGVLLAKRAKKEVGRSSSTEVRRAIHLVSSDWICE 679
QY 355 SFGRIRKSTCL 364
DB 680 -----KNPCL 684

RESULT 3
QY 065552 PRELIMINARY: PRT: 367 AA.

```

```

AC 065552;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE HYPOTHETICAL 39.6 KDA PROTEIN.
GN F6118.110 OR AT4G30980.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxId=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M.; Benes V.; Rechmann S.; Borkova D.; Ansoerge W.; Hohelsel J.;
RA Mewes H.W.; Mayer K.; Schueller C.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benes V.; Rechmann S.; Borkova D.; Ansoerge W.; Mewes H.W.; Lemcke K.;
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL022198; CAI18195.1; -.
DR EMBL: AL161578; CAB79816.1; -.
DR MENDEL: 29120; Arabid.3244; 29120.
DR INTERPRO: IPR001092; -.
DR PFM: PF00010; HLH; 1.
DR KW Hypothetical protein.
SQ SEQUENCE 367 AA; 39613 MW; 80A5E4E8F4ECE6DA CRC64;

```

```

Query Match 5.8%; Score 121; DB 10; Length 367;
Best Local Similarity 24.9%; Pred. No. 0.11;
Matches 45; Conservative 36; Mismatches 82; Indels 18; Gaps 5;

QY 140 HSKERLRERKRYCCEDRLTLPYVGRKNDASVLEATVDYKIRKISPAVMAQIT 199
DB 198 HSIERLRERKRYCCEDRLTLPYVGRKNDASVLEATVDYKIRKISPAVMAQIT 255
QY 200 FALQSNMRCKKQTPIELSLPTVMAORENSVMSTYSPERG--LQFLNTGNCSTPD 257
DB 256 GAASASQISQEDAGSHENTSSSGEAKMTEHYAKLMEDKMSANQYLGK--GLCLMPI 313
QY 258 AESSLDEAVRPSSASANAIDGPYKTHISSAALSL---NSLHVRYSKYTPSYDATAVT 314
DB 314 SLATTISTATCPSSRSPFVKDTEVPPLSPNLSTIIVANGSSSLVTVK-----DAPSV 364
QY 315 T 315
DB 365 S 365

RESULT 4
QY 094771 PRELIMINARY: PRT: 452 AA.
AC 094771;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE TRANSCRIPTION FACTOR-LIKE 5.
GN TCF5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Maruyama O.; Katagiri T.; Nishimori H.; Miki Y.; Ueno A.; Nakamura Y.;
RT "Cloning, mapping, and genomic organization of THL1, a novel human
testis-specific gene containing a basic Helix-loop-Helix motif.";

```


OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96267138; PubMed=9602032;
 RA Schweinfest C.W., Graber M.W., Henderson K.W., Papas T.S., Baron P.L.,
 Watson D.K.;
 RT "Cloning and sequence analysis of Hsp90alpha DeltaN, a new member of
 the hsp90 gene family.";
 RL Biochim. Biophys. Acta 1398:16-24(1998).
 DR EMBL: AF028832; AAC25497.1; -.
 DR INTERPRO: IPR001404; -.
 DR PFAM: PF00183; HSP90; 1.
 SQ SEQUENCE 539 AA; 63251 MW; F10A801D002B6547 CRC64;

Query Match 5.4%; Score 113.5; DB 4; Length 539;
 Best Local Similarity 21.8%; Pred. No. 0.72;
 Matches 81; Conservative 58; Mismatches 141; Indels 91; Gaps 18;

QY 6 VPSSIAEELAIKILRFGRK---KNTSLFEVFI--PENFGGICSGHMDIAL---TPP 57
 DB 201 LPLNLSREMLQOSKILKVRKNLVKCKLELFTLEADKENYKFTYQFSKNIKLGIHDS 260
 QY 58 LTMKSNVYKWTCTPSTVKTENATGPEELG---PLQSYSEHLGYFPDGL----- 108
 DB 261 QNRKKLSLELRYT-----SASGDEWVSLKDYCTRMKNQKHIIYITGETKDQVA 310
 QY 109 -FACSESLNGLNGLEL-----NASLSEFEKNKKIS-----LLHSSKEKLRE 149
 DB 311 NSAFYERLIR-HGLEVIYMEPIDEXVOQLKEFEKGKTLVSYTKEGLELPEDEEEKKOE 369
 QY 150 RIKYCCOELRTLLPYKGRKNDASVLEATVDYVKIRREKI--SPVMAQITEALQSNMF 208
 DB 370 EKKTFENLCKIM-----KDLKKVKEVYVSNLVSPCCIVSTIGWYANMER 419
 QY 209 CKQQTPIELSLPGVMAQRENSVNSTYSPERGLOFLNTCWNCGSTPDSESLDEAVY 268
 DB 420 IMKAQ-----ALRDNSTMGVMAKKHLEI-----NPD--HSIIFTELQ 455
 QY 269 PS-SSASENAID---PKTHISSAALSUN--SLHTVRYYS--KTPSYDATATVYNOVI 319
 DB 456 KMEADKNDKSVADVILLVETALLSGFSLDPQTHANRIYKMIKLGIDDDPTADPT 515
 QY 320 SIHLPSAMPV 330
 DB 516 SAAVTEEMPL 526

RESULT 8
 Q9SOS3 PRELIMINARY; PRT; 311 AA.
 AC Q9SOS3;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE PUTATIVE DNA-BINDING PROTEIN.
 GN T6L1.19.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
 RA Ronning C.M., Koo H., Fujii C.Y., Uterback T.R., Barnstead M.E.,
 RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome I BAC T6L1 genomic sequence.";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC011665; AAF07356.1; -.
 DR INTERPRO: IPR001092; -.
 DR PFAM: PF00010; HLH; 1.

KW DNA-binding.
 SQ SEQUENCE 311 AA; 34487 MW; CC72A040A1A2E427 CRC64;

Query Match 5.4%; Score 113; DB 10; Length 311;
 Best Local Similarity 20.1%; Pred. No. 0.37;
 Matches 57; Conservative 46; Mismatches 125; Indels 56; Gaps 8;

QY 99 EHLGTFPDDLPACSSSLNGLNGLELNASLSEFEKNKKISLHSSKEKLRRERIKCCQL 158
 DB 9 EEDYGEEDFNSKRGPSSNTTVHNSRDSKENDKSAIRSKHSYEQRRRSKINERFQIL 68
 QY 159 RTLLPYVGRKNDASVLEATVDYVKIRREKIS-----PVMQITEALQSNMFC 209
 DB 69 RELINSE-QKRDYASFLLEVVDYQYLOEKYQKTEGSPKSGOPEPTLTPRNHMRVQ 127
 QY 210 KKQQTPIELS-----LPGTVMAQRENSVST---YSPERGLOFLNTCWNCGSTPD 257
 DB 128 SLGNHPVAINNNGSGPIFPCKF---EDNTVSTPAIIAEQPIESDKARALITGSI-E 183
 QY 258 AESSIDEAVRP-----SSASENAGDPYKTHISSAALSUNSLHTVRYYS 303
 DB 184 SQPELDDGLPLPILPMPVQEQANECPATSDGQSNDLVIEGTTISSAVSHELLS 243
 QY 304 KTPSYDATAV-----TNQISIHLPAMPVVS 331
 DB 244 SUTQALQNAIGDLSQAKLSVQIDLGRANQGLTHEPSSKNPLS 287

RESULT 9
 Q9ZUG9 PRELIMINARY; PRT; 350 AA.
 AC Q9ZUG9;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE F27D4.17 PROTEIN.
 GN F27D4.17.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
 RA Shen M., Ronning C.M., Fraser C.M., Somerville C.R., Venter J.C.;
 RT "Arabidopsis thaliana chromosome II BAC F27D4 genomic sequence.";
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC005967; AAD03387.1; -.
 DR MENDEL: 38842; Arabid.3244.38842.
 DR INTERPRO: IPR001092; -.
 DR PFAM: PF00010; HLH; 1.
 SQ SEQUENCE 350 AA; 36522 MW; 8438D16596C636B9 CRC64;

Query Match 5.3%; Score 112; DB 10; Length 350;
 Best Local Similarity 37.5%; Pred. No. 0.52;
 Matches 27; Conservative 16; Mismatches 17; Indels 12; Gaps 2;

QY 140 HSKREKLRRERIKYCCOELRTLLPYVGRKNDASVLEATVDYVKIRREKI----- 190
 DB 149 HSIAERLRERIERAKMKALQELVP--NGNKTDKASMLDEIIDYKFLDQYKVLMSRLG 206
 QY 191 -SPVMAQITEA 201
 DB 207 GAASVSQISEA 218

RESULT 10
 Q9M270 PRELIMINARY; PRT; 359 AA.
 ID Q9M270

```

AC Q9M270;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE HYPONETICAL 41.2 KDA PROTEIN.
GN F21P14.120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Choise N., Robert C., Brottier P., Mincker P., Catolico L.,
RA Attiguenave F., Saurin W., Weissenbach J., Mewes H.W., Lemcke K.,
RA Meyer K.F.X., Queller F., Salanoubat M.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Arabidopsis sequencing project;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: A1138642; CAB71902.1; -.
KW Hypothetical protein.
SQ SEQUENCE 359 AA; 41158 MW; BCATFAT95EEDBDBA CRC64;

Query Match
Best Local Similarity 22.3%; Score 112; DB 10; Length 359;
Matches 65; Conservative 4; Mismatches 107; Indels 76; Gaps 12;

QY 82 NATGPEELGPL-----QSYSEHLGYFPDPLFCAGSESLRNGLELNASISEFNKKI 136
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 125 NMTLPSTSPSLSHSRKRRKRNHL--LPQEMTREKRKRRTKPSKNNEEI-----ENQRI 178
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 137 SLHSSSEKRLREIKKCCEDRLTLPLPYGKRNDAASYLEATVDYKRIEKSIPVMA 196
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 179 N-HIAERRRRRMMHNSLRLLPSPYIQRGDQASIVGAINYK-----VLE 227
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 197 QTEALSNMFKCKKQOTPIELSLPGTVMAQRENSVMSTYSPERGLOFLTNTMNGCSTP 256
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 228 QIIOSLESQR-----TQOQNSSEV-----VENALNLT-----SGTSSN 261
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 257 DAESSLDEAVRPSSAS-----ENAIQDPYKTHISSAALSLSLH--TVRY 301
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 262 DLMFTLEDQICPIKIEATVIONHVSLLKQVCEKKGQLKGIISLEKLTALVLIHNTTSS 321
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 302 YSKVTPSYDATAVTNNQNIHSLPAMPVSSFSGLTALGMARALHPTV 352
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 322 HSSVSYSFNLKQMEDF---CDLESA-----DETTAAVHRIFDIPPTI 359
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
Q9XEFO PRELIMINARY; PRT; 254 AA.
AC Q9XEFO;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HYPONETICAL 28.4 KDA PROTEIN.
GN T07M07.8 OR AT2G40200.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA MDLINE-99225673; Pubmed-10207155;
RA Wang M.L., Belmonte S., Kim U., Dolan M., Morris J.W., Goodman H.M.;
RT "A cluster of ABA-regulated genes on Arabidopsis thaliana BAC
T07M07."
RL Genome Res. 9:325-333(1999).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Unayen L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana."
RL Nature 402:761-768(1999).
DR EMBL: AF085279; AAD25935.1; -.
DR EMBL: AC018721; AAP18734.1; -.
DR INTERPRO: IPR001092; -.
DR PFM: PF00010; HLH; 1.
KW Hypothetical protein.
SQ SEQUENCE 254 AA; 28414 MW; EE3BCE353EE52619 CRC64;

Query Match
Best Local Similarity 22.2%; Score 111; DB 10; Length 254;
Matches 69; Conservative 47; Mismatches 109; Indels 86; Gaps 13;

QY 79 KENATGPEELGPLQMSYE-----HIGYFPDPLFCAGSESLRNGLELNASISFEK 132
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 9 KWSPTTPYVWSWSLQESSSDPMNRPNLGF-----SSSFGN-FPADCVGIEK 59
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 133 NKRIISLHSEKRLRRRIKCCEDRLTLPLPYGKRNDAASYLEATVDYKRIEKSIP 192
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 60 AESLSRSHRLAEKRDRINSHLTALKVLP--NSDKLDAAALATVIEQYKELQKAA- 116
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 193 AYVAQITREALQSNMFKCKKQOTPIELSLPGTVMAQRENSVMSTYSPERGLOFLNT 248
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 117 -----ESPFIQDLF-----TEADVTVQPTISDFESNTTII 149
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 249 -----CWNGCSTPDAESSLDEAVRPSSASENAIGDPTYTHISSAALSLSLHVTYIS 303
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 150 FKASFC---CE--DQPAISEIIRVLKIQLET-----IOAELISVGGRRIRNFIL 195
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 304 KVPESYDAT--AVTNNQNIHSLPAMPVSSFSGLTALGMARALHPTVNSGFRKS 361
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 196 KDSNCNETTINIASAKALKOSLCSALNRTSSSTITS-----SVC-----RIRS 239
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 362 TCLKFTLSTY 372
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 240 KRWKFLSHX 250
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
Q9ZPY8 PRELIMINARY; PRT; 662 AA.
AC Q9ZPY8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE F13A10.4 PROTEIN.
GN F13A10.4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.,
RA Carreira A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,
RA Fraser C.M., Venter J.C.;
RT "Arabidopsis thaliana chromosome II BAC F13A10 genomic sequence."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC006418; AAD20162.1; -.
DR INTERPRO: IPR001092; -.

```

```
DR INTERPRO: IPR003015; -.
DR PFAM: PF00010; HLH; 1.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN 1.
DR SEQUENCE 662 AA; 74546 MF; A6C7ECC03AABF53F CRC64
```

Query Match	5.3%;	Score 110;	DB 10;	Length 662;
Best Local Similarity	21.0%;	Pred. No. 1.8;		
Matches	83;	Conservative	58;	Mismatches 154;
				Indels 100;
				Gaps 177;

[illegible]

RESULT	13		
090US4			
ID	090US4	PRELIMINARY;	PRT; 339 AA.
AC	090US4;		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, last sequence update)		
DT	01-OCT-2000 (TREMBLrel. 15, last annotation update)		
DE	BASIC HELIX-LOOP-HELIX FACTOR 1 (BASIC-HELIX-LOOP-HELIX PROTEIN) (HES		
DE	RELATED REPRESSOR PROTEIN 1 HERP1).		
GN	CHP1 OR HRT2 OR HEY2.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
NR	[1]		
RP	SEQUENCE FROM N.A.		
RA	Chin M.-T., Maemura K., Fukumoto S., Jain M., Layne M., Watanabe M.,		
RA	Hsieh C.-M., Lee M.-E.;		
RT	"Cardiovascular basic helix-loop-helix factor 1, a novel		
RT	transcriptional repressor expressed preferentially in the developing		
RL	rat and adult cardiovascular system.";		
RL	J. Biol. Chem. 0:0-0(2000).		
NR	[2]		
RP	SEQUENCE FROM N.A.		
RA	Nakagawa O., Nakagawa M., Richardson J.A., Olson E.N., Srivastava D.,		
RT	"HRT1, HRT2 and HRT3: A New Subclass of bHLH Transcription Factors		
RT	Marking Specific Cardiac, Somitic and Pharyngeal Arch Segments.";		
RL	Dev. Biol. 216:72-84(1999).		
NR	[3]		
RP	SEQUENCE FROM N.A.		
RA	Steidl C., Lehmestier C., Klamt B., Meier M., Nanda I., Schmid M.,		
RA	Gebauer M.;		
RT	"Characterisation of the human and mouse Hey genes: cloning, mapping		
RT	and mutation screening of a new bHLH gene family.";		

RL submitted (Jan-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Leimaster C., Externbrink A., Kiant B., Gessler M.;
 RT "Hcy genes: a novel subfamily of hairy- and Enhancer of split related
 RL genes specifically expressed during mouse embryogenesis."; *Mech. Dev.* 85:173-177(1999).

RA SEQUENCE FROM N.A.
 RP Iso T., Sartorelli V., Polatz C., Chen T., Suvov H.M., Chung G.,
 RA Wu H.-X., Kedes L., Hamamori Y.; -
 RT "The HERP Repressors, Novel Partners for HES/ hairy in Notch
 RT Signaling."; -
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF173902; AAF20174.1; -
 DR EMBL; AF172287; AAF15446.1; -
 DR EMBL; AJ271867; CAB71346.1; -
 DR EMBL; AF232240; AAF37298.1; -
 DR INTERPRO: IPR00104; -
 DR INTERPRO: IPR001092; -
 DR INTERPRO: IPR001993; -
 DR INTERPRO: IPR003015; -
 DR PFAM: PF00010; HTH_1.
 DR PRINTS: PR00308; ANTI-FREZE1.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN.1.
 DR PROSITE: PS00215; MITOCH_CARRIER; UNKNOWN.1.
 SQ SEQUENCE 339 AA: 35873 MW: 1055690DDE50444 CRC64;

Query Match	5.28;	Score 109;	DB 11;	Length 339;
Best Local Similarity	21.78;	Pred. No. 0.86;		
Matches	61;	Conservative	42;	Mismatches 108;
			Indels	70;
			Gaps	9

```

OY 87 EEGLPLOSSEHGYPITDLPFACSESLRNGNLELNASLSEEEKKKKISLHSSNEKL 146
Db 16 ETDGSSNNPFGH-----ATSSWMSNSPTTTSQIMARKKRGIT-----EKR 59
OY 147 RREIRKYCCEDLRTLLPYVKGRKND-----ASVLEATVYVKYIREKISPAVM---AQIT 195
Db 60 RDRDRINNSLELRVLPTAFEGKQSAKLEKAEKELIOMTVMDHKKMLQATGKGYPADHALAT 115
OY 200 EALSSNMFC---KKOOTPIELSLPGVMAQRENSVASTSPERGQIFLNTCMNCST 255
Db 120 DFMSTGFECLTEVARYLSSVEGLDPSDPLKVRVLVSHLSTASOR-----164
OY 256 PDAESSLEAVRVSSSSASENAIDPYK-----THISAALSLSLHTVRYKSKVTPSYD 310
Db 165 -----EAAVVMSSMAHHNHPRLPHRHHMAAFHHLPTLALDORGLHT-----SSTTCRL 212
OY 311 ATATVTONISTHLPSPMPVVSFSLGALLSGMARALHTIP 351
Db 213 STS-----SEVPSAGSALLTATFAHNASALRMP 242

```

RESULT	14		
094895			
ID	094895	PRELIMINARY;	PRT; 1708 AA.
AC	094895;		
DT	01-MAY-1999	(TRMBLrel, 10, Created)	
DT	01-MAY-1999	(TRMBLrel, 10, Last sequence update)	
DT	01-MAY-2000	(TRMBLrel, 11, Last annotation update)	
DE	KIAA0803 PROTEIN (FRAGMENT).		
GN	KIAA0803.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID:9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=BRAIN;		
RX	MEDLINE=99087487; PubMed=9872452;		
RA	Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,		
RA	Kotani H., Nomura N., Ohara O.;		

RT "Prediction of the coding sequences of unidentified human genes. XI.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro."
 RL DNA Res. 5:277-286(1998).
 DR EMBL: AB018346; BAA34523.1; -
 FT NON_TER 1
 SQ SEQUENCE 1708 AA; 197004 MW; A4232490DCB1E229 CRC64;

Query Match 5.2%; Score 109; DB 4; Length 1708;
 Best Local Similarity 20.6%; Pred. No. 7.9; Mismatches 94; Indels 54; Gaps 9;
 Matches 50; Conservative 45;

QY 13 EELEAI-----KLIRFGKK--KNTHSLFVFIIPENFGKIGSGHMDIALTEPLTME 61
 DB 148 EEIQLENEVIEKIQOELANIGQKTSMAHSL-----SEADSLKHQLDVYIAEKLALE 200
 QY 62 KMSNVVKYITTCPSNTVKE---NATGPELGLPLQRSYSEHLGTFP----- 105
 DB 201 QOVETANEEMTFMKNVKLTETNFKMNQLTQELFSLKRRRESVEKIQSIPIENSVNVAIDHLS 260
 QY 106 -----TDLFACSESLRNGNGLEINASTSEFEKKKISILHSSKEKLRRE-----RIK 152
 DB 261 KDKPELEVLTEDAKS---LENQYFKSFEEGKSGIINLETRLQLLESTVSADKDELT 317
 QY 153 YCCEDQLRTLIPYVKGKNDASVLEATVDYKYIREKISPAVMAOI-TEALOSNMRFCKK 211
 DB 318 QCYKQIKDM---QEOGOFETEMLOKKIYVNLQKIVEEKVAAALVSOILEAVQEVAKFCOD 374
 QY 212 QQT 214
 DB 375 NOT 377

RESULT 15

QY6B8 PRELIMINARY: PRT: 1773 AA.

AC QY6B8;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)
 DE SPLICED VARIANT AKAP350 (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN NCBI_TaxID=9606;
 RP SEQUENCE FROM N.A.
 RC TISSUE=LUNG;
 RA Milgram S.L., Goldenring J.R., Schmidt P.H.;
 RT "AKAP350: A multiply spliced family of proteins with centrosomal
 RT association."
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF091711; AAD39719.1; -
 FT NON_TER 1
 SQ SEQUENCE 1773 AA; 204414 MW; 092A3174725D9262 CRC64;

Query Match 5.2%; Score 109; DB 4; Length 1773;
 Best Local Similarity 20.6%; Pred. No. 8.4; Mismatches 94; Indels 54; Gaps 9;
 Matches 50; Conservative 45;

QY 13 EELEAI-----KLIRFGKK--KNTHSLFVFIIPENFGKIGSGHMDIALTEPLTME 61
 DB 195 EEIQLENEVIEKIQOELANIGQKTSMAHSL-----SEADSLKHQLDVYIAEKLALE 247
 QY 62 KMSNVVKYITTCPSNTVKE---NATGPELGLPLQRSYSEHLGTFP----- 105
 DB 248 QOVETANEEMTFMKNVKLTETNFKMNQLTQELFSLKRRRESVEKIQSIPIENSVNVAIDHLS 307
 QY 106 -----TDLFACSESLRNGNGLEINASTSEFEKKKISILHSSKEKLRRE-----RIK 152
 DB 308 KDKPELEVLTEDAKS---LENQYFKSFEEGKSGIINLETRLQLLESTVSADKDELT 364

QY 153 YCCEDQLRTLIPYVKGKNDASVLEATVDYKYIREKISPAVMAOI-TEALOSNMRFCKK 211
 DB 365 QCYKQIKDM---QEOGOFETEMLOKKIYVNLQKIVEEKVAAALVSOILEAVQEVAKFCOD 421
 QY 212 QQT 214
 DB 422 NOT 424

Search completed: January 19, 2001, 10:50:14
 Job time: 28350 sec

[The main body of the page contains a large, faint, and mostly illegible document. A single horizontal line is visible across the middle of the page.]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2001, 23:07:12 ; Search time 2530.03 Seconds
(without alignments)
4304.523 Million cell updates/sec

Title: US-09-389-000-1

Perfect score: 2128
Sequence: 1 gaccggggggcggttggggt.....ccacattatcgaataaaaaa 2128

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenBank: *
1: gb_bai: *
2: gb_baz: *
3: gb_om: *
4: gb_ov: *
5: gb_ph: *
6: gb_p11: *
7: gb_p12: *
8: gb_p13: *
9: gb_p14: *
10: gb_p15: *
11: gb_p16: *
12: gb_p17: *
13: gb_p18: *
14: gb_p19: *
15: gb_p20: *
16: gb_p21: *
17: gb_p22: *
18: gb_p23: *
19: gb_p24: *
20: gb_p25: *
21: gb_p26: *
22: gb_p27: *
23: gb_p28: *
24: gb_p29: *
25: gb_p30: *
26: gb_p31: *
27: gb_p32: *
28: gb_p33: *
29: gb_p34: *
30: gb_p35: *
31: gb_p36: *
32: gb_p37: *
33: gb_p38: *
34: gb_p39: *
35: gb_p40: *
36: gb_p41: *
37: gb_p42: *
38: gb_p43: *
39: gb_p44: *
40: gb_p45: *
41: gb_p46: *
42: gb_p47: *
43: gb_p48: *

44: em_hg2: *
45: em_hg3: *
46: em_hg4: *
47: gb_p13: *
48: gb_p14: *
49: gb_p15: *
50: gb_p16: *
51: gb_p17: *
52: gb_p18: *
53: gb_p19: *
54: gb_p20: *
55: gb_p21: *
56: gb_p22: *
57: gb_p23: *
58: gb_p24: *
59: gb_p25: *
60: gb_p26: *
61: gb_p27: *
62: gb_p28: *
63: gb_p29: *
64: gb_p30: *
65: gb_p31: *
66: gb_p32: *
67: gb_p33: *
68: gb_p34: *
69: gb_p35: *
70: gb_p36: *
71: gb_p37: *
72: gb_p38: *
73: gb_p39: *
74: gb_p40: *
75: gb_p41: *
76: gb_p42: *
77: gb_p43: *
78: gb_p44: *
79: gb_p45: *
80: gb_p46: *
81: gb_p47: *
82: gb_p48: *
83: gb_p49: *
84: gb_p50: *
85: gb_p51: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1528.6	71.8	2147	37 AK000456	AK000456 Homo sapi
2	337.4	15.9	15888	67 AL160392	AL160392 Homo sapi
3	202	9.3	133864	37 AL139377	AL139377 Human DNA
4	58.8	2.8	7218	81 I6494	I6494 Sequence 14
5	57	2.7	318	77 HS47C6R	HS47C6R Sequence 14
6	48	2.3	2069	21 E10125	E10125 DNA encodin
7	48	2.3	3399	21 E10126	E10126 DNA encodin
8	45	2.1	132513	48 AC004520	AC004520 Homo sapi
9	45	2.1	188563	50 AC022764	AC022764 Homo sapi
10	44	2.1	201010	54 AC034122	AC034122 Mus muscu
11	43.8	2.1	132491	53 AC026784	AC026784 Homo sapi
12	43.8	2.1	162162	39 AC012432	AC012432 Homo sapi
13	43.8	2.1	162397	39 AC012536	AC012536 Homo sapi
14	43.8	2.1	174057	58 AC073127	AC073127 Homo sapi
15	43.6	2.0	19985	56 PCU43145	PCU43145 Plasmodium
16	43.2	2.0	204207	50 AC023147	AC023147 Homo sapi
17	42.8	2.0	10030	1 AE001134	AE001134 Borrelia
18	42.8	2.0	141948	29 AC008413	AC008413 Homo sapi
19	42.8	2.0	181195	49 AC021230	AC021230 Homo sapi
20	42.8	2.0	197748	69 AL390840	AL390840 Homo sapi
21	42.4	2.0	226798	50 AC022235	AC022235 Mus muscu

C	22	42	2.0	21202	17	PFC034008
C	23	42	2.0	92611	9	AC0065398
C	24	42	2.0	319241	58	AC073132
C	25	41.6	2.0	163916	70	AP002413
C	26	41.6	2.0	167584	70	AP001843
C	27	41.6	2.0	168681	51	AC023670
C	28	41.4	1.9	92493	7	AT2327
C	29	41.4	1.9	109584	10	AC010461
C	30	41.4	1.9	110000	30	AC0085572_3
C	31	41.4	1.9	126059	30	AC008862
C	32	41.4	1.9	161915	10	AC023169
C	33	41.2	1.9	111122	7	ATF151
C	34	41.2	1.9	119119	6	AC008970
C	35	41.2	1.9	122543	9	AC028214
C	36	41	1.9	99411	8	AC005207
C	37	41	1.9	145338	54	AC036154
C	38	40.8	1.9	71082	9	AC007393
C	39	40.8	1.9	111187	70	AP001801
C	40	40.8	1.9	152831	49	AC021743
C	41	40.8	1.9	167498	70	AP000750
C	42	40.8	1.9	171111	70	AP000787
C	43	40.8	1.9	174742	41	AC016896
C	44	40.8	1.9	183539	53	AC026869
C	45	40.8	1.9	196348	39	AC01452

ALIGNMENTS

[illegible][illegible]

FEATURES
SOURCE

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="KATO III"
/cell_type="signal-ring cell carcinoma"
/clone="KAT0575"
/clone_1b="KAT"
/note="cloning vector pME18FL3"
14. .1291

```

Oy	701	cagagagccttggagatgatcatctcaacatggtctcttgaagtgtcctttcac	760
Db	180	CAGAGGCGCTTTGGAGATGATTCGATTTCACTGGTTCTCTGAAGGCGCTTCTCAC	239
Oy	761	taagtcgcsagagcttggaagccatcaagttaatgatttggcaaaaagaatacac	820
Db	240	TAACTGCCGGAGGAGCTGGAGGCCATCAAGTTAATTAAGATTGTGGCAAAAAGAAATATCAC	299
Oy	821	attacagtttgttttataatacccgaaattttaaaggttgatcttcacagagcatgaa	880
Db	300	ATTACTGTTTGTTTTATATATCCCTGAAAATTTTAAAGGTGTATTTCAAGGCACTGGAA	359
Oy	881	tgagtaattgctttaaactgaaaccactgacaaatggaanaaaatgagtaatgtgtgtaataact	940
Db	360	TGATATGCTTTAACTGAACCACTGACATGGAATAAATAAGATGATGGTAAATACT	419
Oy	941	ggaaacaacatgccctcaaaacactggttaagaactggaagccaactggtgcttgaaagacttg	1000
Db	420	GGAACAAATGTCCCTCAAACACTGTTTAAACATGAAAACCCAACTGGGCTGAAGAACTTG	479
Oy	1001	gattgccctcgcsagaggtctccacagagaaacctggaattcttctaactgactatcttg	1060
Db	480	GATTGCCCTCGCAGAGGTCTTACAGGAACACTGGGATTTTCTCTACTATCATTTTG	539
Oy	1061	ccgtctctgaactcttaaggaatgacatggaatgggcttgaaatlaaagtcttcgttgtcaagt	1120
Db	540	CCCTCTCTGATCTTTAAAGGAATGGCAATGGGCTGGAATTAATCTTCGTGTGACAGT	599
Oy	1121	tcgagaaaaaacaagaatctctctcttcattcaagcagaagaaaaaacttaagaagagaa	1180
Db	600	TCGAAAAACAAAAAATCTCTCTTCTTCAATTCAGCAGAGAAAACTTAAGAGGGAAA	659
Oy	1181	gaatcaaatatgtctgtgagcaagctgcgtaactctctgcgcatgtlaaagaggagaaga	1240
Db	660	GAATCAATATTTGGTGTGAGCAGACTCGTACACTCTTCCGTATGTAAAAAGGAAGAAGA	719
Oy	1241	atgtgtgggcttcaagtctcttgaggcaaacggttgattatgtgaaatataatccggsgaaaa	1300
Db	720	ATGATGGGGCTTCAAGTCTTGAGGCAACAGTTGATTATGTGAATATATCCGGGGAAAA	779
Oy	1301	tctctcagccgttaatgtggcccaagataagaagaacttaagagcaaatgaagtttga	1360
Db	780	TCTCTCCAGCGCTTAATGCGCCCAATTCAGAAACAACTTCAAGCAACAATGAGCTTTTGTA	839
Oy	1361	agaaacaacaacaccccatgagctgtctctccagagcactgtcatgtggcaacagcgggaaa	1420

Db 54781 ACAATGGCCCTGACCTACACTCTTTCTCAAAAAATATTCACAAATTATGAAA 54722
OY 2124 aaa 2126
Db 54721 AAA 54719

RESULT 3
AL139377/c
LOCUS AL139377 153964 bp DNA PRI 16-SEP-2000
DEFINITION Human DNA sequence from clone RP11-251J8 on chromosome 13, complete
sequence.
ACCESSION AL139377
VERSION AL139377.8 GI:10185469
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 153964)
AUTHORS Phillimore,B.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Sep 18, 2000 this sequence version replaced gi:9943981.
requests: clonerequest@sanger.ac.uk
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-251J8 is from the library RP11-11.1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://Daqpac.med.buffalo.edu/
VECTOR: pBAC3.6
This sequence is the entire insert of clone RP11-251J8 The true
left end of clone RP11-22001 is at 76095 in this sequence. The true
right end of clone RP11-121N13 is at 46542 in this sequence.
FEATURES
source
1..153964
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-251J8"
/clone_11b="RP11-11.1"
complement(4394..4902)
/note="match: GSS: Em:AQ566052"
4913..5434
/note="match: GSS: Em:B56355"
20504..21023
/note="match: GSS: Em:AQ565583"
complement(28424..28906)
/note="match: GSS: Em:AQ519597"
complement(37716..38198)
/note="match: GSS: Em:AQ726251"
49445..49809
/note="match: GSS: Em:AQ017175"

misc_feature complement(53217..53653)
/note="match: GSS: Em:AQ527490"
misc_feature complement(58957..59671)
/note="match: GSS: Em:AQ375951"
65026..66528
/note="match: GSS: Em:AQ036199"
complement(69728..70139)
/note="match: GSS: Em:AQ218718"
72359..72935
/note="match: GSS: Em:AQ386749"
74084..74298
/note="match: GSS: Em:AQ08740"
74131..74254
/note="match: GSS: Em:AQ379168"
complement(78982..79522)
/note="match: GSS: Em:AQ699184"
complement(79085..79496)
/note="match: GSS: Em:AQ267170"
79965..80602
/note="match: GSS: Em:AQ353722"
81464..81824
/note="match: GSS: Em:AQ027170"
88495..88841
/note="match: GSS: Em:AQ088142"
complement(102691..103082)
/note="match: GSS: Em:AQ435358"
complement(102692..103043)
/note="match: GSS: Em:AQ088271"
complement(105513..105949)
/note="match: GSS: Em:AQ434114"
105972..106395
/note="match: GSS: Em:AQ402870"
106283..106790
/note="match: GSS: Em:AQ282655"
114150..114790
/note="match: GSS: Em:AQ317196"
complement(121221..121649)
/note="match: GSS: Em:B56354"
complement(121255..121609)
/note="match: GSS: Em:AQ466633"
123153..123847
/note="match: GSS: Em:AQ527227"
132103..132500
/note="match: GSS: Em:B55736"
132118..132493
/note="match: GSS: Em:AQ021461"
complement(132120..132493)
/note="match: GSS: Em:AQ344420"
132181..132416
/note="match: GSS: Em:A49391"
complement(132230..132493)
/note="match: GSS: Em:B94003"
132260..132541
/note="match: GSS: Em:AQ627818"
complement(132296..132496)
/note="match: GSS: Em:AQ617736"
complement(132323..132495)
/note="match: GSS: Em:AQ021494"
complement(132355..132502)
/note="match: GSS: Em:AQ034473"
complement(132358..132495)
/note="match: GSS: Em:AQ314706"
complement(132379..132499)
/note="match: GSS: Em:AQ313377 Em:AQ553345"
133586..133962
/note="match: GSS: Em:AQ207066"
147570..147573
/note="152 element removed from here."
BASE COUNT 48904 a 31432 c 28348 g 45280 t
ORIGIN
Query Match 9.5% Score 202; DB 37; Length 153964;

Best Local Similarity 100.0%; Pred. No. 3,3e-38;
Matches 202; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 73506 GACCGGGGGCGGTGGGGTTCACCCCTCGTGGCTTGTGGTGGCCCTTAA 73447

1 gaccggggggttggttggttcaccgctcgtcgtactggtcttgggtggtccctaa 60
|||||
|||||

Db 73446 TGTCTTGTCTCTAAGTCTGTAGGGGAAAGACCGGAGTCTCTGGCTTACACTATG 73387

61 tgtcttgccttaagtgcttgagggaagagcgggaggtccttgctgactacatg 120
|||||
|||||

Db 73386 AAGGAGAGAGAACTACAACTCGACGGTGTGACCAACCGCTGAACAGCAGTTG 73327

121 aaggaagagaacacacacttcagcgggtgtgagaccacacgcctgaaacagcgttg 180
|||||
|||||

Db 73326 CTGGAAGAAGTCCGCAAGAAGT 73305

181 ctggaagaagtcgcgaagaagt 202
|||||
|||||

RESULT 4
166494/c 166494 7218 bp DNA PAT 28-DEC-1997
LOCUS 166494
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dornier, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
source 1. 7218
location/Qualifiers
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match 2.8%; Score 58.8; DB 81; Length 7218;
Best Local Similarity 5.2%; Pred. No. 0.0012;
Matches 21; Conservative 224; Mismatches 161; Indels 0; Gaps 0;

Db 82 gaggggaaagcgcggaggtcctgcctgacacatgaaggaagagaactacaac 141
|||||
|||||

Db 1429 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1370

142 ttgcaggtgtgagcaaccacgcctgaacagcagttgctggaagaagtcgcgaagaag 201
|||||
|||||

Db 1369 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1310

202 taggaagagaaacccgttatgagacacactgaataatgagtagaagaaagca 261
|||||
|||||

Db 1309 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1250

262 aattgttatcgaagaaagtgaaataatcgaatgaagaaacttcagatagacttc 321
|||||
|||||

Db 1249 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1190

322 ttcttctgtctatgaagaaatccagtggaatccttaacacattacttaacagct 381
|||||
|||||

Db 1189 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1130

382 agaagaagaagagacacttgaaagcaagtgaaataactatgcaactaaactgaag 441
|||||
|||||

Db 1129 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1070

442 agaataaagcttacagaagatcaacaatgaagcgcgtacatc 487
|||||
|||||

Db 1069 RRRATCGCAAGCTCCCTGACCTCGAGCCAGCTCGAATTAATTC 1024

RESULT 5
HS47C6R 318 bp DNA PRI 17-OCT-1995
LOCUS HS47C6R/c
DEFINITION H.sapiens Cpg Island DNA genomic MseI fragment, clone 47c6, reverse
read cpq47c6.rta.
ACCESSION 255522.1 GI:1021563
VERSION 255522.1
KEYWORDS Cpg Island; genomic MseI fragment.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 318)
AUTHORS MacDonald, M., Huckle, E., Wilkinson, P. and Micklem, G.
TITLE Direct Submision
JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk
REFERENCE 2 (bases 1 to 318)
AUTHORS Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.
TITLE Purification of Cpg islands using a methylated DNA binding column
JOURNAL Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE 94282070
COMMENT Vector: pGEM-5zf(-)
Clones are available from the UK MRC Human Genome Mapping Project
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
http://www.hgmp.mrc.ac.uk/ for details
or contact: biolhelp@hgmp.mrc.ac.uk.
FEATURES
source 1. 318
location/Qualifiers
BASE COUNT 57 a 102 c 105 g 54 t
ORIGIN

Query Match 2.7%; Score 57; DB 77; Length 318;
Best Local Similarity 100.0%; Pred. No. 0.0032;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 57 GACCGGGGGCGGTGGGGTTCACCCCTCGTGGCTTGTGGTGGCCCT 1

1 gaccggggggttggttggttcaccgctcgtcgtactggtcttgggtggtccct 57
|||||
|||||

Db 57 GACCGGGGGCGGTGGGGTTCACCCCTCGTGGCTTGTGGTGGCCCT 1

RESULT 6
E10125 standard; RNA; UNC; 2069 BP.
ID E10125
XX
AC E10125;
XX
SV E10125.1
XX
DT 08-OCT-1997 (Rel. 52, Created)
DT 02-SEP-2000 (Rel. 65, Last updated, Version 2)
DE DNA encoding an immunogenicity protein.
DE JP 1995284392-A/1.
KW
XX
XX
OS unidentified
OC unclassified.
XX
XX
RN [1]
RP 1-2069
RA Doi H., Nagakuchi Y., Tanaka Y., Fujisaki Y.,
RT "GENE CLONE MANIFESTING CHICKEN LEUCOCYTOZOON IMMUNOGENIC PROTEIN AND GENE
RECOMBINANT VACCINE AGAINST CHICKEN LEUCOCYTOZOON";

Db 2467 GAAGAAAAAGAGAGTACATGAGAGAAAAAGAGTACATGAGAGAA 2526
 Oy 467 aacatgaacgcctacactagctgaatgctcagg 506
 Db 2527 GAAAGATACATGAGAGAGAAAAAGTACATG 2566

RESULT 8
 AC004520/c LOCUS AC004520 132513 bp DNA PRI 03-FEB-2000
 DEFINITION Homo sapiens BAC clone CTB-119C2 from 7p15, complete sequence.
 AC004520
 VERSION AC004520.1 GI:3004572
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 132513)
 Dubuque,T., Smith,A., Elliott,G. and Harmon,G.
 The sequence of Homo sapiens BAC clone CTB-119C2
 Unpublished
 2 (bases 1 to 132513)
 Waterston,R.
 Direct Submission
 Submitted (01-APR-1998) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 3 (bases 1 to 132513)
 Waterston,R.
 Direct Submission
 Submitted (03-FEB-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: saplens@watson.wustl.edu
 Summary Statistics
 Center project name: H_RG119C02

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see
 http://www.nhgri.nih.gov/DIR/CTB/CHR/, send
 mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
 Clone CTB-119C2 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (http://www.resgen.com).
 VECTOR: pBelbac11
 Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
 The actual start of this clone is at base position 1 of CTB-119C2; actual end is at 132513 of CTB-119C2. The orientation of this clone is unknown.

This clone contains STS SWS2790 (NID:g1133585).

FEATURES
 Location/Qualifiers
 1. 132513

source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7p15"
 /clone="CTB-119C2"
 /clone_1id="CTB-978SK-B"
 /complement(92..580)
 /note="match to EST AA53630 (NID:g2292107) vk57e11.r1"
 /complement(1168..1437)
 /note="match to EST H99041 (NID:g1123709) yx11h12.s1"
 /complement(1202..1437)
 /note="match to EST R66707 (NID:g839345) y133d08.s1"
 /complement(1272..1437)
 /note="match to EST AA036731 (NID:g1509988) zk30c06.s1"
 /complement(1293..1437)
 /note="match to EST AA745905 (NID:g2785891) ob02h02.s1"
 /complement(1305..1437)
 /note="match to EST AA701656 (NID:g2704821) z143f09.s1"
 /complement(1322..1437)
 /note="match to EST N27564 (NID:g1142045) yy01h07.s1"
 /complement(1349..1437)
 /note="match to EST AA678000 (NID:g2858522) z114f08.s1"
 /complement(1370..1437)
 /note="match to EST AA788753 (NID:g2848873) ah30g06.s1"
 1380..1437
 /note="match to EST (NID:g1434444)"
 /complement(1386..1437)
 /note="match to EST AA150090 (NID:g1721330) z142a06.s1"
 /complement(1396..1437)
 /note="match to EST AA701656 (NID:g2704821) z143f09.s1"
 /complement(1396..1437)
 /note="match to EST R66707 (NID:g839345) y133d08.s1"
 /complement(1396..1437)
 /note="match to EST N27564 (NID:g1142045) yy01h07.s1"
 /complement(1396..1437)
 /note="match to EST AA678000 (NID:g2858522) z114f08.s1"
 /complement(1396..1437)
 /note="match to EST H99041 (NID:g1123709) yx11h12.s1"
 /complement(1396..1437)
 /note="match to EST AA150090 (NID:g1721330) z142a06.s1"
 /complement(1396..1437)
 /note="match to EST AA036731 (NID:g1509988) zk30c06.s1"
 /complement(1396..1437)
 /note="match to EST AA745905 (NID:g2785891) ob02h02.s1"
 1963..2262
 /note="match to EST (NID:g1434444)"
 /complement(1963..2301)
 /note="match to EST AA788753 (NID:g2848873) ah30g06.s1"
 3282..3373
 /rpt_family="MALR"
 4620..4819
 /rpt_family="MER2-type"
 4925..5022
 /rpt_family="Alu"
 6135..6571
 /note="match to EST W76339 (NID:g1386789) zd60f11.s1"
 6149..6531
 /note="match to EST H21401 (NID:g890096) y132g06.s1"
 /complement(6641..39925)
 /gene="WUGSC:H_RG119C02.1"
 /complement(join(6641..7891,8640..8723,14302..14481,39356..39925))
 /gene="WUGSC:H_RG119C02.1"
 /note="similar to NFE2-related transcription factors;

```

similar to I48694 (PID:g2137676); H_RG119C02.1"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC09039.1"
/db_xref="GI:3004573"
/translation="MKHLKRWMSAGGGLLHTLLSLIAGLRVDLDLTLPPETLLOD
ELLPFGRRASSAYALSPSPASGGMKRRAGHLHKGKRELDPAAPPEQOLLRVRLALVPR
VPRISVDAMLVHSAVASGADDEAHGGLGAAASSTGAGASVDGSOAVQGGGDFRAA
RSGPLDAGEEERKAPAPTAQVPDAGGASSEENGVLREKHEAVDSSOHEMERVSAQ
KENSILQDDDDENKIAERKDEMEKTEKRENRHNGDTFSFLEDLQOLSSQSPENS
LEGISLIDPLPESISDGMNSAHYVNSQAIQDVNHEALILCPNTEFRDPTAR
TSOSEPPLQINSHTNPEQTLPGNTLGFSLSPVNHMMTLTSCDLLDIDINDEI
NLMGATEDNPPPIVSOQLFDEPDSGSLSDSSNNNTSVTKSNSHSDGGAIGYC
TDHSSSHHDEGAAGGTYPEPSKLDHDSQSDHGDUTFEQVHEHNTYHLQPTAPE
STSEPPMPGKSKQIRSRYLEDTDRNLSDERAKALHLPESVDIYGVPMFNSML
SRVYLDLOVSLRIDIRGRKNVAQNRKRRLDIILLEDVQNLQAKKETLRQO
AQCNAINIMKOKLHDLYHDIFSRRLRDDGGRVNNHVALQCTHDGSLIYKELVAS
GHKKTOKRKR"
complement(7777..7893)
/ gene="WUGSC:H_RG119C02.1"
/ note="match to EST AA135857 (NID:g1697161) z025h07.r1"
misc_feature
complement(7781..7897)
/ gene="WUGSC:H_RG119C02.1"
/ note="match to EST AA149705 (NID:g1720647) zn99f04.r1"
misc_feature
complement(7798..7897)
/ gene="WUGSC:H_RG119C02.1"
/ note="match to EST AA099770 (NID:g1646643) z181a01.r1"
repeat_region
complement(8593..8725)
/ rpt_family="L1"
misc_feature
complement(8602..8862)
/ gene="WUGSC:H_RG119C02.1"
/ note="match to EST 144351 (NID:g1048879)"
repeat_region
complement(14301..14483)
/ rpt_family="Alu"
9338..9614
/ rpt_family="Alu"
9792..10082
/ rpt_family="Alu"
11148..11429
/ rpt_family="Alu"
12696..12873
/ rpt_family="Alu"
12908..13196
/ rpt_family="Alu"
complement(14301..14483)
/ gene="WUGSC:H_RG119C02.1"
/ note="match to EST AA149705 (NID:g1720647) zn99f04.r1"
misc_feature
complement(14302..14483)
/ gene="WUGSC:H_RG119C02.1"
/ note="match to EST AA099770 (NID:g1646643) z181a01.r1"
misc_feature
complement(14302..14483)
/ gene="WUGSC:H_RG119C02.1"
/ note="match to EST AA135857 (NID:g1697161) z025h07.r1"
misc_feature
complement(14302..14483)
/ gene="WUGSC:H_RG119C02.1"
/ note="match to EST AA132584 (NID:g1694136) z020c03.r1"
repeat_region
14796..16547
/ rpt_family="SVA"
16630..16927
/ rpt_family="Alu"
17052..17359

```

```

Query Match 2.1%; Score 45; DB 8; Length 132513;
Best Local Similarity 52.4%; Pred. No. 2.5;
Matches 99; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

```

```

QY 1004 tgccttcagagagctcctacagcaacctggatattcttactatgatctttgacct 1063
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126022 TGACTAAGTTGAGAGCTCTCTCCGTCACCTCAGTAATGTCTCAGATCCAGTTGCC 125963

```

```

QY 1064 gctctgaactttaaggaatgcaatgagcttgatataatgcttcgtgtcagaattcg 1123
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125962 ATGTAGATACCTATTGAAGAAAGCTCTTTGACTGAAATTAACATTAATAACCTATTG 125903

QY 1124 agaaacacaaagaagctctctcttcattcaatcaagaagaaactaagaaggaagaa 1183
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125902 TGGAGCCCAATTAATAAAGCATATTTCATTGGTGCAAACTTCGAAAGACGACATTA 125843

QY 1184 tcaaatatt 1192
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125842 TTCATTATT 125834

RESULT 9
AC022764 188563 bp DNA HTG 16-MAR-2000
LOCUS Homo sapiens chromosome 17 clone RP11-310M15 map 17, WORKING DRAFT
DEFINITION
ACCESSION AC022764.3 GI:7249221
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 188563)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckwith,R., Bida,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeRubeis,K., Dewar,K., Domino,M., Doyle,M., Feneator,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardina,S., Grant,G., Hagos,B., Heathford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,B., Kahn,L., Karasik,A., Klein,J.,
Lander,T., Lechoczky,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGuirk,A., McKernan,K.,
McPheters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olyar,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rotman,D.,
Roy,A., Santos,R., Severi,P., Spencer,B., Strange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,D., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2000 this sequence version replaced gi:16983429.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Project -----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: 15848
Center clone name: 310.M.15
----- Summary Statistics -----
Sequencing vector: M13; W77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 16496 bases at least Q40
Consensus quality: 176314 bases at least Q30
Consensus quality: 180940 bases at least Q20
Insert size: 210000; agarose-fp
Insert size: 184653; sum-of-ontigs
Quality coverage: 2.7 in Q20 bases; agarose-fp
Quality coverage: 3.1 in Q20 bases; sum-of-ontigs

```

NOTE: This is a 'working draft' sequence. It currently consists of 40 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
1 1196: contig of 1196 bp in length
1197 1296: gap of 100 bp
1297 2422: contig of 1126 bp in length
2423 2522: gap of 100 bp
2523 3938: contig of 1316 bp in length
3939 5333: contig of 1395 bp in length
5334 5433: gap of 100 bp
5434 6892: contig of 1459 bp in length
6893 6992: gap of 100 bp
6993 9051: contig of 1959 bp in length
9052 10497: gap of 100 bp
10498 10597: contig of 1446 bp in length
10598 12563: contig of 1966 bp in length
12564 12663: gap of 100 bp
12664 14771: contig of 2108 bp in length
14772 14871: gap of 100 bp
14872 16733: contig of 1862 bp in length
16734 16833: gap of 100 bp
16834 19429: contig of 2596 bp in length
19430 19529: gap of 100 bp
19530 21429: contig of 1900 bp in length
21430 21529: gap of 100 bp
21530 24676: contig of 3147 bp in length
24677 24776: gap of 100 bp
24777 27200: contig of 2424 bp in length
27201 27300: gap of 100 bp
27301 29853: contig of 2553 bp in length
29854 29953: gap of 100 bp
29954 32573: contig of 2620 bp in length
32574 32673: gap of 100 bp
32674 34834: contig of 2161 bp in length
34835 34934: gap of 100 bp
34935 38184: contig of 3250 bp in length
38185 38284: gap of 100 bp
38285 41429: contig of 3145 bp in length
41430 41529: gap of 100 bp
41530 43659: contig of 2130 bp in length
43660 43759: gap of 100 bp
43760 46302: contig of 2543 bp in length
46303 46402: gap of 100 bp
46403 49933: contig of 3531 bp in length
49934 50033: gap of 100 bp
50034 54240: contig of 4207 bp in length
54241 54340: gap of 100 bp
54341 57127: contig of 2787 bp in length
57128 57227: gap of 100 bp
57228 61476: contig of 4249 bp in length
61477 61576: gap of 100 bp
61577 65550: contig of 3974 bp in length
65551 65650: gap of 100 bp
65651 70580: contig of 4930 bp in length
70581 70680: gap of 100 bp
70681 76028: contig of 5348 bp in length
76029 76128: gap of 100 bp
76129 83825: contig of 7697 bp in length
83826 83925: gap of 100 bp
83926 90939: contig of 7014 bp in length
90940 91039: gap of 100 bp
91040 96426: contig of 5387 bp in length
96427 96526: gap of 100 bp
96527 101964: contig of 5438 bp in length
101965 102064: gap of 100 bp

102065 108545: contig of 6481 bp in length
108546 108645: gap of 100 bp
108646 114408: contig of 5843 bp in length
114409 114588: gap of 100 bp
114589 122283: contig of 7659 bp in length
122284 122383: gap of 100 bp
122384 131395: contig of 9012 bp in length
131396 131495: gap of 100 bp
131496 142144: contig of 10649 bp in length
142145 142244: gap of 100 bp
142245 154513: contig of 12269 bp in length
154514 154613: gap of 100 bp
154614 169786: contig of 15173 bp in length
169787 169886: gap of 100 bp
169887 188563: contig of 18677 bp in length.
Location/Qualifiers
1. 188563
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-310M15"
/clone_11b="RP11 Human Male BAC"
1. 1196
/note="assembly_fragment"
1297. 2422
/note="assembly_fragment"
2523. 3838
/note="assembly_fragment"
3939. 5333
/note="assembly_fragment"
5434. 6892
/note="assembly_fragment"
6993. 8951
/note="assembly_fragment"
9052. 10497
/note="assembly_fragment"
10598. 12563
/note="assembly_fragment"
12664. 14771
/note="assembly_fragment"
14872. 16733
/note="assembly_fragment"
16834. 19429
/note="assembly_fragment"
19530. 21429
/note="assembly_fragment"
21530. 24676
/note="assembly_fragment"
24777. 27200
/note="assembly_fragment"
27301. 29853
/note="assembly_fragment"
29954. 32573
/note="assembly_fragment"
32674. 34834
/note="assembly_fragment"
34935. 38184
/note="assembly_fragment"
38285. 41429
/note="assembly_fragment"
41530. 43659
/note="assembly_fragment"
43760. 46302
/note="assembly_fragment"
clone_end:17
vector_side:left
46403. 49933
/note="assembly_fragment"

Query Match 2.1%; Score 45; DB 50; Length 188563;
Best Local Similarity 48.3%; Pred. No. 2.5; Mismatches 0; Gaps 0;
Matches 126; Conservative 0; Mismatches 135; Indels 0;

```

QY 238 attaatgatgaactagaagaactgtttatctcaaggagaagtggaataatcca 297
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116423 ATTGAATGCTGGAATTAATTACAGAAAAATCAATACGCGAAAAATACAGAAAGAG 116482
QY 298 tgaactctcctgagatgactcttctatctcgctgctatgaagaaatgcagcagc 357
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116483 ATAGTAATTAACAGAGAAATCAGGTAGAGAGCTGTGAATGACAAATGCAAGAGAAAA 116542
QY 358 cttaacacactactactaagaagctagaagaagaagaagactcttgaagaactgaa 417
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116543 GGAATAATGATTAATAAACACACAGATGAAGAGAAATTTAAACTTTTGAGATTAATGAT 116602
QY 418 atactatgcacttaactgaagaagaagaagcttaccagaagatcaacaatgaacg 477
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116603 TATCTACCCGACACATCCAAACATATCTGACATCTGTGAATACAGAAATGAGATATA 116662
QY 478 ccgtactactactagctgaat 498
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116663 TTCAGAGTACTCTGCACGAAAT 116683

RESULT 10
AC034122 201010 bp DNA HTG 26-MAY-2000
LOCUS Mus musculus clone RP23-169C19, WORKING DRAFT SEQUENCE, 30
DEFINITION unoriented pieces.
ACCESSION AC034122
VERSION AC034122.3 GI:8077093
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 201010)
REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 201010)
REFERENCES Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
Collins,A., Cooke,P., DeRubeis,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,M., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hages,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehotzky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPherson,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rotman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Teste,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
COMMENT Submitted (04-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:7770516.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
JOURNAL Genome Center
TITLE Center: Whitehead Institute/ MIT Center for Genome Research.
COMMENT Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information

```

```

Center project name: L8058
Center clone name: 109_C-19
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 18419 bases at least Q40
Consensus quality: 182258 bases at least Q30
Consensus quality: 195732 bases at least Q20
Insert size: 208000; agarose-fp
Insert size: 198110; sum-of-contigs
Quality coverage: 3.9 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1111: contig of 1111 bp in length
* 1112 1211: gap of 100 bp
* 1212 2404: contig of 1193 bp in length
* 2405 2504: gap of 100 bp
* 2505 3976: contig of 1472 bp in length
* 3977 4076: gap of 100 bp
* 4077 5334: contig of 1258 bp in length
* 5335 5434: gap of 100 bp
* 5435 6696: contig of 1262 bp in length
* 6697 6796: gap of 100 bp
* 6797 9202: contig of 2406 bp in length
* 9203 9302: gap of 100 bp
* 9303 10287: contig of 985 bp in length
* 10288 10387: gap of 100 bp
* 10388 12441: contig of 2054 bp in length
* 12442 12541: gap of 100 bp
* 12542 14480: contig of 1939 bp in length
* 14481 14580: gap of 100 bp
* 14581 16881: contig of 2301 bp in length
* 16882 16981: gap of 100 bp
* 16982 18958: contig of 1977 bp in length
* 18959 19058: gap of 100 bp
* 19059 21985: contig of 2927 bp in length
* 21986 22085: gap of 100 bp
* 22086 24348: contig of 2263 bp in length
* 24349 24448: gap of 100 bp
* 24449 28133: contig of 3685 bp in length
* 28134 28233: gap of 100 bp
* 28234 32277: contig of 4044 bp in length
* 32278 32377: gap of 100 bp
* 32378 38555: contig of 6178 bp in length
* 38556 38655: gap of 100 bp
* 38656 44903: contig of 6248 bp in length
* 44904 45003: gap of 100 bp
* 45004 52544: contig of 7541 bp in length
* 52545 52644: gap of 100 bp
* 52645 61054: contig of 8410 bp in length
* 61055 61154: gap of 100 bp
* 61155 68886: contig of 7732 bp in length
* 68887 68986: gap of 100 bp
* 68987 77233: contig of 8247 bp in length
* 77234 77333: gap of 100 bp
* 77334 86303: contig of 8970 bp in length
* 86304 86403: gap of 100 bp
* 86404 96573: contig of 10170 bp in length
* 96574 96673: gap of 100 bp
* 96674 107775: contig of 11102 bp in length
* 107776 107875: gap of 100 bp
* 107876 119036: contig of 11161 bp in length
* 119037 119136: gap of 100 bp
* 119137 133728: contig of 14592 bp in length

```

```

* 133729 133828: gap of 100 bp
* 133829 145758: contig of 11930 bp in length
* 145759 145858: gap of 100 bp
* 145859 161359: contig of 15501 bp in length
* 161360 161459: gap of 100 bp
* 161460 178068: contig of 16609 bp in length
* 178069 178168: gap of 100 bp
* 178169 201010: contig of 22842 bp in length.
FEATURES
    Location/Qualifiers
        1..201010
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /clone="RP23-169C19"
        /clone.lib="RPC1-23 Female Mouse BAC"
        1..1111
        /note="assembly-fragment"
misc_feature 1212..2404
    /note="assembly-fragment"
misc_feature 2505..3976
    /note="assembly-fragment"
misc_feature 4077..5334
    /note="assembly-fragment"
misc_feature 5435..6696
    /note="assembly-fragment"
misc_feature 6797..9202
    /note="assembly-fragment"
misc_feature 9303..10287
    /note="assembly-fragment"
misc_feature clone_end:T7
    vector_side:left"
misc_feature 10388..12441
    /note="assembly-fragment"
misc_feature 12542..14480
    /note="assembly-fragment"
misc_feature 14581..16881
    /note="assembly-fragment"
misc_feature 16982..18938
    /note="assembly-fragment"
misc_feature 19059..21985
    /note="assembly-fragment"
misc_feature 22086..24348
    /note="assembly-fragment"
misc_feature 24449..28133
    /note="assembly-fragment"
misc_feature 28234..32277
    /note="assembly-fragment"
misc_feature 32378..38555
    /note="assembly-fragment"
misc_feature 38656..44903
    /note="assembly-fragment"
misc_feature 45004..52544
    /note="assembly-fragment"
misc_feature 52645..61054
    /note="assembly-fragment"
misc_feature 61155..68886
    /note="assembly-fragment"
misc_feature clone_end:SP6
    vector_side:left"
misc_feature 68987..77233
    /note="assembly-fragment"
misc_feature 77334..86303
    /note="assembly-fragment"
misc_feature 86404..96573
    /note="assembly-fragment"
misc_feature 96674..107775
    /note="assembly-fragment"
misc_feature 107876..119036
    /note="assembly-fragment"
misc_feature 119137..133728
    /note="assembly-fragment"
misc_feature 133829..145758
    /note="assembly-fragment"
misc_feature 145859..161359
    /note="assembly-fragment"

```

```

Query Match 2.1%; Score 44; DB 54; Length 201010;
Best Local Similarity 50.5%; Pred. No. 4.4;
Matches 107; Conservativity 0; Mismatches 105; Indels 0; Gaps 0;
misc_feature 161460..178068
    /note="assembly-fragment"
misc_feature 178169..201010
    /note="assembly-fragment"
Db 123752 AGTTTCCATATTTGTATTAAGAACGAGCCAGCCAGCAGCAAGTCCAGTTGA 123811
QY 850 aatttaagggttgatttcaggagcatggaatgatatgtcttaactgaaccactgaca 909
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123812 ATAAATMAAAAGMAACCCACAGTACATGTTCTTAAACCTAAGAACCTAGAGCAAG 123871
QY 910 atggaagaatgaatgaatgtgttaaatctgacaactgtccctcaaaccttga 969
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 970 actgaacacgcacactggcctgaagaacttg 1001
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123872 GTCMAAGTTGMAAAGCATCTCTGAGCACAAG 123903

```

```

RESULT 11
AC026784
LOCUS AC026784
DEFINITION Homo sapiens chromosome 5 clone CMD-2138014, WORKING DRAFT
SEQUENCE, 12 ordered pieces.
ACCESSION AC026784
VERSION AC026784.3 GI:9964791
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 132491)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 132491)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Sep 2, 2000 this sequence version replaced gi:7711994.
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 680891
Center Clone name: CTRB-H1_2138014
-----
Summary Statistics
Consensus quality: 124042 bases at least Q40
Consensus quality: 129692 bases at least Q30
Consensus quality: 130729 bases at least Q20
Estimated insert size: 132010; agarose-fp estimation
Estimated insert size: 131991; sum-of-ctdigs estimation
Quality coverage: 6.48 in Q20 bases; agarose-fp estimation
Quality coverage: 6.48 in Q20 bases; sum-of-ctdigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced

```


Query Match	Similarity	2.1%	Score 43.8	DB 39	Length 162397
Best Local	Similarity	48.2%	Pred. NC 4.9	Mismatches 132	Conservative 0
Matches	123	Conservative	0	Mismatches 132	Indels 0
Gaps	0	Gaps	0	Gaps	0
OY	214	aaccggtatgagacacacacttgaatlaatgatgacagacagaagcaatgttcatc	273		
Db	129048	AAACTTAAGTATATATATATAAAAAATATAATATAATATAATATAATAAACAT	129107		
OY	274	caaggaagaagtggagaaaaatccatcgtgaaccttcacagatagactatcttatcg	333		
Db	129108	CAATGTTAACTGATAGAAAACATGTATATCTGAAAATAGAGCACAATGGATGAT	129167		
OY	334	ctatgaagcaatgacagctggaatccttcaacacataacttaaacagctgaagaagaaa	393		
Db	129168	AGACCAATAGACATAGAAATTGACCCGAAATAGAAATATTTCACATAGACCCACAA	129227		
OY	394	gaagactcttgaagaatcgaatgaatcactgcacttaacatcgtgaacagaatcaaa	453		
Db	129228	GAGAACTAGAAAGTATGTATATCTGAACTGGAAGAAACAGAAATGGGAAAA	129287		
OY	454	ttaccagaagaatcaa	468		
Db	129288	ATTAAGAAATGAAATGAA	129302		

174127 174057 bp DNA

13 unordered pieces.

HTG 02-OCT-2000

WORKING DRAFT

~~Chordata; Cranialata; Vertebrata; Euteleostomi
Imates; Catarrhini; Homnidae; Homo.~~

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0645N11
----- Summary Statistics -----
Sequencing vector: M13; 95%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 95% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 168198 bases at least Q40
Consensus quality: 170189 bases at least Q30
Consensus quality: 171312 bases at least Q20
Insert size: 146000; agarose-fp
Insert size: 172857; sum-of-contigs
Quality coverage: 5.67 in Q20 bases; agarose-fp
Quality coverage: 5.64 in Q20 bases; sum-of-contigs

```

* NOTE: THIS is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	4200:	contig of 4200 bp in length
*	4201	4300:	gap of unknown length
*	4301	8951:	contig of 4651 bp in length
*	8952	9051:	gap of unknown length
*	9053	14534:	contig of 5463 bp in length
*	14535	14634:	gap of unknown length
*	14635	23344:	contig of 8710 bp in length
*	23345	23444:	gap of unknown length
*	23445	33995:	contig of 10551 bp in length
*	33996	34095:	gap of unknown length
*	34096	44270:	contig of 10175 bp in length
*	44271	44370:	gap of unknown length
*	44371	62317:	contig of 17947 bp in length
*	62318	62417:	gap of unknown length
*	62418	90640:	contig of 28223 bp in length
*	90641	90740:	gap of unknown length
*	90741	119137:	contig of 28397 bp in length
*	119138	119237:	gap of unknown length
*	119238	16616:	contig of 45582 bp in length
*	16616	166713:	gap of unknown length
*	166713	167120:	contig of 2893 bp in length
*	167121	167713:	gap of unknown length
*	167713	171819:	contig of 4007 bp in length
*	171820	171819:	gap of unknown length
*	171820	174057:	contig of 2238 bp in length

URES	Location/Qualifiers
source	1. 174057 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="7" /clone="RP11-645N11" 1. 4200
misc_feature	/note="assembly_name:Contig10" 4301. 8951
misc_feature	/note="assembly_name:Contig11" 9052. 14534
misc_feature	/note="assembly_name:Contig12" 9053. 14534

Query Match	2.1%	Score	43.8	DB	58	Length	174057
Best Local Similarity	53.9%	Pred	NO.4.9				
Matches	90	Conservative	0	Mismatches	77	Indels	0
						Gaps	0

RESULT 15
PCU43145

VERSION U43145.1 GI:1151157
KEYWORDS .

ORGANISM *Plasmodium chabaudi*;
Eukaryota; Alveolata; Apicomplexa; Haemosporida; *Plasmodium*

AUTHORS	TITLE	JOURNAL	REFERENCE
Werner, E.B., Holder, A.A., Aszodi, A. and Taylor, W.R.	A novel 11-residue coiled-coil motif predicts a histidine zipper	Protein Pept. Lett. 3 (2), 139-146 (1996)	2 (bases 1 to 9985)

AUTHORS	Werner, E. B., Taylor, W. R. and Holder, A. A.
TITLE	A Plasmodium chabaudi protein contains a repetitive region with a predicted spectrin-like structure
JOURNAL	Mol. Biochem. Parasitol. 94 (2), 185-196 (1998)
WORD COUNT	100

MEDLINE 98418/b5
REFERENCE 3 (bases 1 to 9985)

Submitted (12-DEC-1995) Ekkehard B. Werner, Parasitology, National
 Direct Submission
 AUTHORS
 TITLE
 JOURNAL

Institute for Medical Research
1AA, UK

FEATURES

Source

```
/organism="Plasmodium chabaudi"
/strain="96V"
```

```

/specific_host="Thamnomys rutilians"
/db_xref="taxon:5825"
/lab_host="Mus musculus"
/dev_stage="asexual blood stage"
2158. .7977
CDS

```

```
repeat_region      4294.  .5385  
                    /rpt_type=tandem  
BASE COUNT      4487 a      766 c      1644 g      3088 t  
ORIGIN
```

Query Match	2.0%;	Score 43.6;	DB 56;	Length 9985;
Best Local Similarity	47.7%;	Pred. NO. 5.5;		
Matches 127;	Conservative	0;	Mismatches 139;	Indels 0;
				Gaps 0;

```
Oy 185 aagaagtcgcgaagaagtagtgaatggaataacccgtaatgagacacaacttgaat 244
    |||||
Db 6451 AATATATTAAATATAAAACGTTGAAGATTAAGACAGCGTATGATTTAAATGTGAAAAAAT 6510
```

QY 245 gataaactagaanaagcaaatgtttatctcaagagaaagtggaaaaatccatggaanc 304
 ||||| ||| ||||| ||||| ||| ||| ||||| |||
 Db 6511 AATATCTACAGAAAGAAATTTCTATCTAAAGAGCAATATAAAAGATTTGTAAGCA 657

305 tcttcagatagactatcttctatctggtctatgaacgaatgccagtgcgaatccttaac 364

Db 6571 TATGCTGAAGAATTAAATAGAAAAATAGCTATGATGAAGTAAAGTAAATAATTAACTAAC 663

Q7 365 acattacttaacagctagaaagaaagaagactcttgaagatcaagtgaataactat 424

Db 6631 GATTGAGGAATTAAAAATAAGAACACAGAAGGTGAGAGACAAATTGCTGAGTTAAT 6699

Tue Jan 23 10:54:39 2001

us-09-389-000-1.rge

Page 16

Db 6691 AACTGAGAGATATTAAAGAAAAA 6716

Search completed: January 19, 2001, 02:57:41
Job time: 13829 sec

Tue Jan 23 10:54:39

Seq primer: Forward.

FEATURES
Source Location/Qualifiers
1..617
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE Resequences, MAGE"
/note="Vector: pBluescriptSKm"

BASE COUNT 211 a 116 c 127 g 163 t
ORIGIN

alignment_scores:

Quality: 291.00 Length: 59
Ratio: 4.932 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.305

alignment_block:

US-09-389-000-2 x AW961069 ..

Align seg 1/1 to: AW961069 from: 1 to: 617

1 MetValLeuLeuLysValProSerSerLeuSerAlaGluGluLeuAla 17
|||||
439 ATGCTCTCTTGAAGTGCCTTCTTCACTAGTCCAGAGGCTGGAGC 488
17 aileLysLeuLeuArpPheGlyLysLysAsnThrHisSerLeuPheV 34
|||||
489 CACCAAGTAAATAGATTGGCAAAAGAAAATACACATTCACTGTTG 538
34 aLpheLeuLeuProGluAsnPhelLysGlyLysSerGlyHisGlyMet 50
|||||
539 TTTTATTAATCCCTGAAATTTTAAAGTTGTATTCAGGGCATGATG 588
51 AspLeuAlaLeuThrGluProLeuThr 59
|||||
589 GATATTGCTTAACTGAACCACTGACA 615

seq_name: gb_est74.BE665112

seq_documentation_block:

LOCUS BE665112 503 bp mRNA EST 08-SEP-2000
DEFINITION 153097 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE665112
VERSION BE665112.1 GI:10025321
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus

REFERENCE

AUTHORS Smith T.P.L., Casas E., Stone R.T., Heaton M.P., Grosse W.M., Bennett G.A., Fahrenkrug S.C., Freking B.A., Rohrer G.A., Laegreid W.M. and Keele J.W.
Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle
Unpublished (2000)
Contact: Smith TP
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov

JOURNAL COMMENT

PCR Primers
FORWARD: AGAACAAGTATGACACAT
BACKWARD: GTTTCACAGTACGACG
Plate: 65 row: 6 column: 20
Seg primer: ATTAGGTACACATATAG.
Location/Qualifiers
1..503

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="MARC 4BOV"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from day 20 and day 40 embryos."

alignment_scores:

Quality: 258.50 Length: 90
Ratio: 3.541 Gaps: 3
Percent Similarity: 81.111 Percent Identity: 63.333

alignment_block:

US-09-389-000-2 x BE665112 ..

Align seg 1/1 to: BE665112 from: 1 to: 503

318 AsnLeuSerLeuHisLeuProSerAlaMetProProValSer SerPheS 334
|||||
3 AACACTCAGTGCCTTCCACACAGCTGACCCAAAGTGTCAGATTCT 52
334 erLeuGlyThrAlaLeuLeuGlyTTPAlaArGArGAlaLeuHisLePro 350
|||||
53 TCCTCAGACATGCATTCGTGTGGCCAGGTGTGCACAGCACACCCCA 102
351 ThrValCysAsnSerPheGlyArgGlyLeuSerThrCysLeuLysPheTh 367
|||||
103 ACTGCTGCACAGATTGGCGCTATTAATACACATATTGAAATGCAC 152
367 rLeu.SerThrThrTyrrTPAlaGlnPheAspAsnLeuGlyLysValGlu 383
|||||
153 TCTAGACC...ACCTATTGGGTGCGCTTGACATACAGGAAGAATGGA 199
384 GlnArgMetLeuLeuLysAlaProProLysAspLeuLeuSerLysGluLe 400
|||||
200 CAAGAATGATCTTGAAGCTACACCCAGACCTGATATCAAAAGGTT 249
400 uAlaTrpPheGlyPhe 405
|||||
250 TGCATTGTATGCGCTC 265

seq_name: gb_est51.AW961078

seq_documentation_block:

LOCUS AW961078 624 bp mRNA EST 01-JUN-2000
DEFINITION EST373045 MAGE Resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION AW961078
VERSION AW961078.1 GI:8150657
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 624)
AUTHORS Hegde P., Qi R., Abernathy K., Dharap S., Gaspard R., Gay C., Holt I.E., Saeed A.I., Sharov V., Lee N.H., Yeatman T.J. and Quackenbush J.

TITLE Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray
Unpublished (2000)
CONTACT: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@igr.org
Plate: 152
Seq primer: forward.

FEATURES
source
Location/Qualifiers
1. 624
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE Resequences, MAGF"
/note="Vector: pBluescriptSKm"

BASE COUNT 209 a 117 c 128 g 170 t
ORIGIN

Alignment_scores:
Quality: 181.00 Length: 65
Ratio: 3.175 Gaps: 5
Percent Similarity: 87.692 Percent Identity: 78.462

Alignment_block:
US-09-389-000-2 x AW961078 ..

Align seg 1/1 to: AW961078 from: 1 to: 624

1 MetValLeuLeuLysValProSerSerLeuSerAlaGluGluLeuGln 17
|||||
439 ATGGTCTCTTGAAGCTGCGCTTTTACCACTGCGAGAGCTGGAAC 488
17 alleLysLeuLeuArg.PheGlyLysLysAsnThr.HisSerLeu 33
|||||
489 CATCAAGTAAATTAGAAATTTGGCAAAAGAAATTCACAAATCTCTT 538
33 eValPheLeuLeuProGluAsnPheLysGlySer.GlyHisGly 49
|||||
539 TGTTTTAAATCCCTGAATAATTTAAAGTGTGTTTTCAGAGGCGCTGA 588
50 MetAspLeuLeuThr.GluProLeuThrMet 60
|||||
589 AATGATTTGTTTACTTGAACCCCTGACAAAG 624
seq_name: gb_gss3:AQ196491

seq_documentation_block:
LOCUS AQ196491 452 bp DNA GSS 16-SEP-1998
DEFINITION CIT-HSP-2383N20.TF CIT-HSP Homo sapiens genomic clone 2383N20, DNA
sequence.
ACCESSION AQ196491
VERSION AQ196491.1 GI:3603853
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,U.C.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
JOURNAL Unpublished (1998)
COMMENT Other-GSS: CIT-HSP-2383N20.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/Bac_end_search/Bac_end_search.html.
Seq primer: M13-21
Class: BAC ends

FEATURES
source
Location/Qualifiers
1. 452
/organism="Homo sapiens"
/db_xref="taxon:9606"

FEATURES
source
Location/Qualifiers
1. 354
/organism="Homo sapiens"
/db_xref="taxon:9606"

Alignment_scores:
Quality: 159.00 Length: 59
Ratio: 3.533 Gaps: 1
Percent Similarity: 76.271 Percent Identity: 57.627

Alignment_block:
US-09-389-000-2 x AQ196491 ..

Align seg 1/1 to: AQ196491 from: 1 to: 452

198 lIeThrGluAlaLeuGlnSerAsnMetArgPheCysLysGlnGln 214
|||||
191 ATTACAGACGACTTCACAGACATGAGTTTGTAGAAACAAAC 240
214 PProLeuLeuSerLeuProGlyThrValMetAlaGlnArg..... 228
|||||
241 ACCCATAGCTGTCTCTCCACGACTGTCTATGCGACGCGTATGATA 290
229GluAsnSerValMetSerThrTyrSerProGluArgGly 241
|||||
291 AGGCGTTTGGAAATCGAAGTGTGATATTTACTCTCCCATGTTGAT 340
242 LeuGlnPheLeuThrAsnThrCysTrp 250
|||||
341 GCTAATATGTACACAAACTGCTAT 367
seq_name: gb_gss24:B92720

seq_documentation_block:
LOCUS B92720 354 bp DNA GSS 25-JUN-1998
DEFINITION CIT-HSP-2170M15.TF CIT-HSP Homo sapiens genomic clone 2170M15, DNA
sequence.
ACCESSION B92720
VERSION B92720.1 GI:2975057
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,U.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished (1998)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/Bac_end_search/Bac_end_search.html
Seq primer: M13-21;
Class: BAC ends

```

/clone="2170M15"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/notes="vector: pBelobAC11. Site_1: HindIII; Site_2: HindIII"
HindIII
BASE COUNT      93 a      70 c      67 g      124 t
ORIGIN

alignment_scores:
  Quality:      146.00      Length:      30
  Ratio:        4.867      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 96.667

alignment_block:
us-09-389-000-2 x B92720 ..

Align seg 1/1 to: B92720 from: 1 to: 354

198 11ethrglualeuglnserasmetargphecyslystysglnghn 214
|||||
202 ATTAGACAGACACTTCAGACACATGAGTTTGTAGAAACATCAAC 251
|||||
214 rProilegluleuserleuproglythrvalmetagl 227
|||||
252 ACCCATGAGCTGTCTCTCCAGACGACTGATGCACAG 291

seq_name: gb_est38:AV674451

seq_documentation_block:
LOCUS      AV674451      647 bp      mRNA      EST      05-OCT-2000
DEFINITION AV674451 Nori Satoh unpublished cDNA library Clona intestinalis
CDNA clone c1b15K14 5', mRNA sequence.
ACCESSION  AV674451
VERSION     AV674451.1 GI:10112450
KEYWORDS   EST.
SOURCE      Clona intestinalis.
ORGANISM   Clona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
            Phlebobranchia; Clonidae; Clona.
REFERENCE  1 (bases 1 to 647)
AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-I,T.
TITLE      Expressed genes in clona intestinalis
JOURNAL    Unpublished (2000)
COMMENT    Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-Ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES
  source
    1..647
    /organism="Clona intestinalis"
    /db_xref="taxon:7719"
    /clone="c1b15K14"
    /clone_lib="Nori Satoh unpublished cDNA library"
    /tissue_type="whole animal"
    /dev_stage="tailbud"

BASE COUNT      229 a      137 c      112 g      169 t
ORIGIN

alignment_scores:
  Quality:      136.50      Length:      227
  Ratio:        1.241      Gaps:      7
Percent Similarity: 48.458      Percent Identity: 24.229

alignment_block:
us-09-389-000-2 x AV674451 ..

Align seg 1/1 to: AV674451 from: 1 to: 647

```

```

144 glulysleuargargluarglyleustyrtyrcyscysglnleuargth 160
|||||
44 GAAGAGTCGACGAGAAAGCTATTAACAATTCGTGGACAATCAGCAG 93
|||||
160 rleuleuiprottyrvalysgcllyarglysasnaaspalaalaser 177
|||||
94 TTGTTCGCCGCAATCAGTAACATTAAGAAAGACATGGCATCTTGCTGG 143
|||||
177 lualethrvalaspyrvalysrtyrilearglulysileser..... 191
|||
144 AACAAAGCTTGATACATATTTATGACACAGATGATGAAGAC 193
|||||
192 ...Proalavalmetaglinallethrghualaleuglnserasmetar 207
|||||
194 AACGACACTGTGTTAAATAAGCTATATGTCATACCGACAAAACAGA 243
|||||
207 gphecyslysglnghnhrProilegluleuserleuproglythr 224
|||||
244 AGAAATGAAAATTTTCGAAAATCTTGAAAAGAAAACATTTCAAGTACA 293
|||||
224 almetaglinalrgluanser..... 231
|||
294 CAAGTCAAAAGATCTTAATCTTTCGACACAGCTGTTCTGGGGTCAT 343
|||||
232 valmetserthrtyrserprogluargglyleuglnphleuthrnsn 248
|||||
344 TCCTTTCAAACTACACTCCACAA..... 367
|||||
248 rCysTTPasnrglycserthrProaspalagluserSerleuasplua 265
|||||
367 ..... 367
|||||
265 lavalargvalProserSerAlasergluasnalaileglyaspro 281
|||||
368 .....CCATCATCA.....AATGCAGATGTGATCCG 397
|||||
282 tyrlsThrhisIleSerAla.....AlaIleSerle 293
|||||
398 TTATTGAAACATCATCAGCAGCACCATCTTATTGATGATCATTAATTCG 447
|||||
293 uasnSerleuHisThrValargtyrtyrSerlyval..... 305
|||||
448 AAGATGGGAATCACCAGTCATCATCAGAAAGATCAAGATTCATGT 497
|||||
306 .....ThrProserTyraaspalaalThrnsnIle 319
|||||
498 TTGCCTCACAACCTCTTTCAGTGTGTCATCATCAGCAGCTGATTA 547
|||||
320 SerIleHisleu..... 323
|||||
548 TCATATCCCATGCAAACTACAGGAGTACATCAAGATCTATCTATCTCC 597
|||||
324 ProserAlametProValaserSerPhe 333
|||||
598 TCCAGCTGCAAAATGTGCCAGTGAACAATTT 628
|||||

seq_name: gb_gss28:CNS05GA7

seq_documentation_block:
LOCUS      CNS05GA7      1015 bp      DNA      GSS      26-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence SP6 end of clone
025108 of library B from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION  AL336040
VERSION     AL336040.1 GI:8229798
KEYWORDS   GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
ORGANISM   Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Eurypterygii; Ctenosquamata; Acanthomorpha; Eucanthomorpha;
            Holacanthopterygii; Acanthopterygii; Percormorpha;

```

REFERENCE 1 (bases 1 to 1015)
 Tetradontiformes: Tetradontoidae; Tetradontidae; Tetraodon.
 Roest-Crolius, H., Jaillon, O., Dasilva, C., Fzames, C., Fisher, C.,
 Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
 Weissenbach, J.
 TITLE Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1015)
 Roest-Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
 Bernot, A., Fzames, C., Wincker, P., Brotlier, P., Quetier, F.,
 Saurin, W. and Weissenbach, J.
 TITLE Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 1015)
 Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 COMMENT This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/tetraodon.
 FEATURES
 source
 1. 1015
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone_id="B"
 /note="Genoscope sequence ID : COAB025BE04B1-end : SP6"
 BASE COUNT 231 a 293 c 271 g 204 t 16 others
 ORIGIN
 Alignment scores:
 Quality: 121.00 Length: 264
 Ratio: 0.858 Gaps: 14
 Percent Similarity: 49.648 Percent Identity: 26.408
 Alignment block:
 US-09-389-000-2 x CNS05GA7 ..
 Align seg 1/1 to: CNS05GA7 from: 1 to: 1015
 66 valvalvlytyrtyrthrcyspro..... 74
 167 ctgtcagatgctgtggaaacacagctgctgctcatccacatgcttgg 216
 75serasnthrvallvtyrthgluasnalathrglyproglu 89
 217 atgctaattgcccccgctataaaanacgaa...gcaccttccccagagc 263
 89 euglyleuproleuglnarvserlyrserglnhileuglytyrphpro 105
 264 cagacctacgctgcagccacagagaaagcctgacttca.....cct 306
 106 thraspluuphealacysserglusleu..... 115
 307 gcagacatgctcctgctcccaacagctactcctcatccgacatcagat 356
 116argasnlyasnlyleuglu 123
 357 gtccgataaagannaccatcagactaagtagacagacagatgttcgcag 406
 123 euasnalserleusergluphegluylasnllyslleserleu 139
 407 agccgcgtgctgctgctgctgctcaccctgctgctgctgctgctg 456
 140 hisserlyr.....gluylsleuarggluargllelysty 153
 457 cagctgagaaacccgttgtagaaaagatgcacagatccgatcaacag 506
 153 rcycysgluyluargtlhrleu.....protyrvalysglu 168

507 ctgcattcatcagctcagctcagctcagctcagctcagctcagctcag 556
 168 rglvysasn.....aspalaaservalleuglualethrvallasp 181
 557 aaccacactcagctgagagagccgacatcctgagatgac..... 600
 182 tyrvallvtyrlylearggluylsleerpro alavalmethlagini 198
 601 ..gtcacttctcagcagacagctgagccggccctctggcg.... 643
 198 lethglualaleuglnaserametarphnecyslyslglnlthr 214
 644gcantacagccagagcgcttcgacactcctgaggaatcgcc 685
 215 proileugluserleuproglythrvalmethlaglnaargluasnse 231
 686 ccccttccctctcagagggctccgcaca.....gaggtccc 720
 231 rvalmetserthrvserprogluarglyleuglnpheurthrasnt 248
 721 ggtcactctcctgcagggcctccacagca.....cagcagctccctcag 764
 248 hrctytrpasnglycysserthrproaspalagluserserleuasp 264
 765 cccacagagggcgagctcagaccacagcaccacccctcagagacagc 814
 265 alavalarvalproserseralaserlualsnalaleglyasp 281
 815agcagcagcagcagcagggcccggtgagagcc 849
 281 ctyrlsyrthrlsleerlalaaleuserleuasnserleuhist 298
 850 ctgc.....tagacctgagaccacagcaga 875
 298 hrvalargtyrtyrserlyrvalthrproserlyrvalalaval 314
 876 ctgtcagacgtgagggagctgacacccctcctatgattgagannacgtg 925
 seq_name: gb_est49:AM830314
 seq_documentation_block:
 LOCUS AM830314 475 bp mRNA EST 18-JUL-2000
 DEFINITION sm25603.y1 gm-cl028 glycine max cdna clone GENOME SYSTEMS CLONE ID:
 gm-cl028-4926 5' similar to TR:022768 022768 HYPOTHEITICAL 33.1 KD
 PROTEIN ;, mRNA sequence.
 ACCESSION AM830314 GI:7924288
 VERSION AM830314
 KEYWORDS EST.
 SOURCE soybean.
 ORGANISM glycine max
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 Fabales; Fabaceae; Papilionoideae; Glycydes.
 1 (bases 1 to 475)
 REFERENCE
 AUTHORS Shoemaker, R., Keln, P., Vodkin, L., Erpelting, J., Corryell, V., Khanna
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
 Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
 Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
 R., Ratter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
 R., Waterston, R. and Wilson, R.
 TITLE Public Soybean EST Project
 JOURNAL Unpublished (1999)
 COMMENT Contact: Shoemaker R./Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)


```

206 targephycyslysglnglnthrproilgluleuuserleuproglyt 223
|||||
501 ...TTTGGCGTAAACTGGCCGAGGCTTAAGTTGACCAAGACCGGACT 546
223 hr...ValmetlaGlInArGluAsnSerValMetSerThrTySer 237
|||||
547 CCTTGCTGACCTGCTCCTGACCAAGAGTAGTTACAGAGCACCCTTCA 593
seq_name: gb_est72:BE548486

seq_documentation_block:
LOCUS BE548486 884 bp mRNA EST 09-AUG-2000
DEFINITION 601076062F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3461918 5',
mRNA sequence.
ACCESSION BE548486
VERSION BE548486.1 GI:9777131
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 884)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/BLML at:
http://image.llnl.gov
Plate: LHAM8458 row: 1 column: 15
High quality sequence stop: 684.
Location/Qualifiers
FEATURES
Source 1. 884
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3461918"
/clone_lib="NIH_MGC_12"
/tissue_type="choriocarcinoma"
/lab_host="DH10B"
/note="Organ: Placenta; Vector: PCMV-SPORT6; Site_1: NciI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.4 kb. Library prepared by Life
Technologies."
BASE COUNT 259 a 191 c 247 g 187 t
ORIGIN
align_scores:
Quality: 119.00 Length: 101
Ratio: 1.776 Gaps: 3
Percent Similarity: 66.337 Percent Identity: 28.723
alignment_block:
US-09-389-000-2 x BE548486 ..
Align seg 1/1 to: BE548486 from: 1 to: 884

140 HisSerSerLysGlnLysLeuArgGluArgIleIleTyrCysCysG1 156
|||||
257 CATACCGAATGGAAGAGATGAGCGCAGAAATCCGATTGCTGTGA 306
156 uGlnLeuArgThrLeuLeuProTyrValLysGlnArgLysAsnAspAlaA 173
|||||
307 TGAGTGAATCTCTAGTGCCTTGTGCAATGCC...GAGATGACAGAG 353
273 laservalLeuGlnAlaThrValAspTyrValLysTyrIleArgGlnLys 189
|||||

```

```

354 COACNACTCTGACGTGAGCACAAGCATTTCTGAAATACATCCAGAAAGA 403
190 lIeSerProAlaValMetAlaGlnIleThrGlnAlaLeuGlnSerAsnMe 206
404 CATGAGATTCCTCTTAAAAAGCAATTTGAGCGTA..... 439
206 targephycyslysglnglnthrproilgluleuuserleuproglyt 223
|||||
440 ...TTTGGCGTAAACTGGCCGAGGCTTAAGCTGACCAAGACCGGACT 485
223 hr...ValmetlaGlInArGluAsnSerValMetSerThrTySerPro 238
|||||
486 CCTTGCTGACCTGCTCCTGACAGGAGTAGTTACAGAGCACCCTGATG 535
239 Gln 239
536 GAG 538
seq_name: gb_est75:BE757087

seq_documentation_block:
LOCUS BE757087 449 bp mRNA EST 15-SEP-2000
DEFINITION 211433 MARC 2B0V Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE757087
VERSION BE757087.1 GI:10171079
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovine; Bos.
REFERENCE 1 (bases 1 to 449)
AUTHORS Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Kottel,G.A., Laegreid
,W.W. and Keele,J.W.
TITLE Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.960904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCAGAGC
Plate: 62 row: 0 column: 24
Seq primer: ATTAGTGTACACTATAG.
Location/Qualifiers
FEATURES
Source 1. 449
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 2B0V"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, thymus,
semitendinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."
BASE COUNT 121 a 108 c 139 g 81 t
ORIGIN
align_scores:
Quality: 116.00 Length: 101
Ratio: 1.731 Gaps: 3
Percent Similarity: 66.337 Percent Identity: 27.723
alignment_block:

```



```

DEFINITION      z187a08.r1 Stratagene colon (#937204) Homo sapiens cDNA clone
ACCESSION      IMAGE:511574 5', mRNA sequence.
VERSION        AA127107
KEYWORDS
SOURCE
ORGANISM       Homo sapiens
                human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS       Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chapellat,B.,
                Chissee,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawke,J.,
                'M., Hultman,M., Kucaba,T., Lacey,M., Le,M., Le,N., Mardis,E., Mc
                'B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
                Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevisan
                Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
                Generation and analysis of 280,000 human expressed sequence tags
                Genome Res. 6 (9), 807-828 (1996)
                97044478
TITLE
JOURNAL        MEDLINE
COMMENT        Contact: Wilson RK
                Washington University School of Medicine
                4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
                Tel: 314 286 1800
                Fax: 314 286 1810
                Email: estewartson.wustl.edu
                This clone is available royalty-free through LNL ; contact the
                IMAGE Consortium (info@image.llnl.gov) for further information.
                Insert Length: 1357 Std Error: 0.00
                Seq primer: -28M13 rev2 from Amersham
                High quality sequence stop: 310.
                Location/Qualifiers
                1..437
                /organism="Homo sapiens"
                /db_xref="GDB:3844069"
                /db_xref="taxon:9606"
                /clone="IMAGE:511574"
                /clone_1fb="Stratagene colon (#937204) "
                /tissue_type="tumor"
                /cell_line="T84 carcinoma cell line"
                /lab_host="SOLR cells (kanamycin resistant)"
                /note="Organ: colon; Vector: pBluescript SK-; Site:1:
                Ecotri; Site_2: XhoI; Cloned unidirectionally. Primer:
                Oligo dT. T-84 colonic epithelial cell line. Average
                insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor
                sequence: 5' GAATTCGGCAGCAG 3' ~3' adaptor sequence: 5
                CTCGACGTTTTTTTTTTTTTTT 3'"
BASE COUNT      123 a 101 c 128 g 83 t 2 others
ORIGIN
alignment_scores:
                Quality: 112.50 Length: 83
                Ratio: 2.045 Gaps: 2
                Percent Similarity: 66.265 Percent Identity: 30.120
alignment_block:
US-09-389-000-2 x AA127107 ..
Align seg 1/1 to: AA127107 from: 1 to: 437
140 HHSerSerIySGlulsEuArGaRgUaRgIleIyStyRcYcYsgI 156
|||||
|||||
134 CATAAACCGAATCGAAAGAGATGAAGAGCCGACAGATCCGCAATTCGTGTA 163
156 uGlnLeuArGThrLeuLeuProTrYValIySGIyArGlyAsnAspIaa 173
184 TGAAGTGAATCTCTTATGTCGGCTTGCAATGCC...GAACATGACAG 230
173 laseRValLeuGlnUaIaThVaIAspTrYValIyStyRleArGULyS 189
|||||
|||||
231 CCACACACTCTGCAGTCGACACAGCAATCTCTGAATCATCAGGAAGA 280
190 lIeSerProAlaValMetAlaGlnIleThrGluAlaLeuGlnSerAsn 206

```

```

281 CATGGAGATTCTCTTAAAAAATTGAGACCGTA..... 316
206 targetCysLysLysGlnGlnThrProIleGluLeuSerLeuProGly 222
    ||||| ||| :|||: |||||
317 ....TTTTGGGGGAAANAAGTCGCCGACNMAAANAAGCTTGACAGACCGGA 361

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2001, 22:26:51 ; Search time 69.17 Seconds
(without alignments)
4958.062 Million cell updates/sec

Title: US-09-389-000-1
2128

Perfect score: 1 gaccggggggcggttggtggtf.....ccacattatgaaaaaaa 2128

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents.NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/PCITUS.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	58.8	2.8	7218	1	US-08-232-463-14
2	40	1.9	3763	1	US-07-792-865D-1
3	40	1.9	7218	1	US-08-232-463-14
4	37.2	1.7	1360	3	US-08-961-083-37
5	34.8	1.6	1425	3	US-09-009-494-5
6	34.2	1.6	1002	3	US-08-576-240-1
7	34.2	1.6	1602	3	US-08-530-950-3
8	34.2	1.6	1602	3	US-08-888-429A-3
9	34.2	1.6	2924	3	US-09-357-073-1
10	34.2	1.6	2943	2	US-08-788-892-1
11	34	1.6	500	3	US-09-141-000-2
c 12	34	1.6	1544	2	US-08-837-593-1
13	33.2	1.6	747	2	US-08-679-765-2
14	33.2	1.6	747	2	US-09-196-525-2
15	32.8	1.5	3095	5	5231168-1
c 16	32.8	1.5	4032	1	US-08-107-748-3
c 17	32.8	1.5	4032	1	US-08-245-809-4
c 18	32.8	1.5	4032	4	PCT-US92-01385-3
c 19	32.6	1.5	51952	4	US-08-947-823-1
c 20	32.6	1.5	72928	3	US-09-009-913-1
c 21	32.4	1.5	2493	3	US-08-945-056-3
c 22	32.4	1.5	3827	1	US-08-170-294-6
c 23	32.4	1.5	3827	2	US-08-664-855-6
c 24	32.4	1.5	3827	2	US-08-718-751-1
c 25	32.4	1.5	3827	2	US-09-049-289-6
c 26	32.4	1.5	7399	2	US-08-418-848A-9
27	32.4	1.5	9709	2	US-08-188-583-5
28	32.4	1.5	9709	3	US-08-388-353-1

29	32.4	1.5	9709	3	US-08-488-551B-1	Sequence 1, Appl
30	32.4	1.5	15581	3	US-08-646-538-35	Sequence 35, Appl
c 31	32.2	1.5	591	1	US-08-090-523-24	Sequence 24, Appl
c 32	32.2	1.5	591	1	US-08-334-639-3	Sequence 3, Appl
c 33	32.2	1.5	591	1	US-08-398-627-24	Sequence 24, Appl
c 34	32.2	1.5	591	1	US-08-406-858-25	Sequence 25, Appl
c 35	32.2	1.5	591	4	PCT-US94-05275-25	Sequence 25, Appl
c 36	32.2	1.5	2431	3	US-08-714-918-15	Sequence 15, Appl
c 37	32.2	1.5	8791	4	PCT-US96-01735-5	Sequence 5, Appl
c 38	32.2	1.5	10564	1	US-08-206-176-5	Sequence 5, Appl
39	31.8	1.5	824	1	US-08-158-353-1	Sequence 1, Appl
40	31.8	1.5	1486	5	551630-3	Patent No. 551630
41	31.6	1.5	2573	3	US-08-714-918-17	Sequence 17, Appl
42	31.6	1.5	2573	3	US-08-714-918-64	Sequence 64, Appl
43	31.4	1.5	2621	2	US-08-553-619B-8	Sequence 8, Appl
c 44	31.4	1.5	5599	2	US-08-477-451-9	Sequence 9, Appl
c 45	31.4	1.5	5599	2	US-08-477-451-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232.463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935.313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 INMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pT9gpt-F1s
US-08-232-463-14
Query Match 2.8%, Score 58.8, DB 1, Length 7218;


```

; TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
; TITLE OF INVENTION: KINASES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,950
; FILING DATE: 19-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/010001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1602 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-530-950-3

Query Match      1.6%; Score 34.2; DB 1; Length 1602;
Best Local Similarity 49.7%; Pred. No. 2.3;
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 371 cttaaacagctagaagaagaagaagactcttgaaagtcagtaactatgacctt 430
DB 1042 CTCGAACAGCTGTGAGAGGAGCCATCCGCACTCCAGCAGCAAGTCTCTGCGAG 1101
QY 431 aaactggaaacaagaatcaaggcttaccagaagaatcaacaatgacgcgtacataccta 490
DB 1102 TTGTGTGACTTTACCTCAGAGTGTCTTAAGAGAAATTCGCAAGACGCGCTACATACCA 1161
QY 491 gctgaatgtctcagaggtcttggtttacatcaagttcttaaaagcaagcagtg 545
DB 1162 GAGCTAATGCAACATCATTTTTCACCCCTACATGAATCCAAAGAGACAGATGTGG 1216

RESULT 8
US-08-888-429A-3
; Sequence 3, Application US/08888429A
; Patent No. 6136596
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Whitmarsh, Alan
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND ONCOPROTEIN-
; TITLE OF INVENTION: ACTIVATED HUMAN PROTEIN KINASE
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

```

; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/888,429A
; FILING DATE: 07-JUL-1997
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/530,950
; FILING DATE: 19-SEP-1995
; APPLICATION NUMBER: 08/446,083
; FILING DATE: 19-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, Peter J.
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/053001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 299354
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1602 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 244...1245
;
US-08-888-429A-3

Query Match      1.6%; Score 34.2; DB 3; Length 1602;
Best Local Similarity 49.7%; Pred. No. 2.3;
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 371 cttaaacagctagaagaagaagaagactcttgaaagtcagtaactatgacctt 430
DB 1042 CTCGAACAGCTGTGAGAGGAGCCATCCGCACTCCAGCAGCAAGTCTCTGCGAG 1101
QY 431 aaactggaaacaagaatcaaggcttaccagaagaatcaacaatgacgcgtacataccta 490
DB 1102 TTGTGTGACTTTACCTCAGAGTGTCTTAAGAGAAATTCGCAAGACGCGCTACATACCA 1161
QY 491 gctgaatgtctcagaggtcttggtttacatcaagttcttaaaagcaagcagtg 545
DB 1162 GAGCTAATGCAACATCATTTTTCACCCCTACATGAATCCAAAGAGACAGATGTGG 1216

RESULT 9
US-09-357-073-1
; Sequence 1, Application US/09357073
; Patent No. 6033910
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; APPLICANT: Lex M. Cowsett
; TITLE OF INVENTION: ANTISENSE MODULATION OF MAP KINASE KINASE 6 EXPRESSION
; FILE REFERENCE: RTS-0086
; CURRENT APPLICATION NUMBER: US/09/357,073
; CURRENT FILING DATE: 1999-07-19
; NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 1
; LENGTH: 2924
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (341)..(1345)
;
US-09-357-073-1

Query Match      1.6%; Score 34.2; DB 3; Length 2924;
Best Local Similarity 49.7%; Pred. No. 3.2;
```



```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2001, 23:11:13 ; Search time 118.2 seconds
(without alignments)
6763.200 Million cell updates/sec

Title: US-09-389-000-1
Perfect score: 2128
Sequence: 1 gaccggggggggttgggtt.....ccacattatgaaaaaaa 2128

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 480022 seqs, 187831343 residues
Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq.36.*
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	2128	100.0	2128	Human PHELIIX CDNA
2	48	2.3	3399	Chicken leucocytos
3	42.8	2.0	11309	Borrelia burgdorfe
4	42.8	2.0	910715	Borrelia burgdorfe
5	41	1.9	11427	Enterococcus faeca
6	40.4	1.9	164976	Methanococcus jann
7	40	1.9	3763	Sequence encoding
8	40	1.9	3763	Merzoite apical-e
9	38.8	1.8	3029	Borrelia burgdorfe
10	38.8	1.8	7953	Staphylococcus aur
11	38.8	1.8	580073	Mycoplasma genital
12	37.8	1.8	6846	Streptococcus pneu

13	37.2	1.7	1360	19	V27342
14	37.2	1.7	1369	20	Z31402
15	37.2	1.7	1969	20	Z34452
16	37.2	1.7	2082	21	A05408
17	37.2	1.7	2110	14	041223
18	37.2	1.7	1265	19	V52267
19	37	1.7	3411	19	V16518
20	37	1.7	3411	20	X83879
21	36.4	1.7	580073	18	T58840
22	35.8	1.7	222	21	Z42477
23	35.8	1.7	910715	20	X20248
24	35.6	1.7	1360	20	X13435
25	35.6	1.7	1377	20	X61486
26	35.6	1.7	1446	20	X61485
27	35.6	1.7	1766	20	Z31962
28	35.6	1.7	2401	20	X20301
29	35.4	1.7	1349	21	A12410
30	35.2	1.7	1206	21	Z58568
31	35	1.6	17646	18	V74420
32	34.8	1.6	1425	19	V45100
33	34.8	1.6	1425	19	V45101
34	34.8	1.6	1425	19	V33272
35	34.8	1.6	1425	20	X56687
36	34.8	1.6	3393	20	X24681
37	34.8	1.6	3400	21	Z51225
38	34.8	1.6	10636	20	X13011
39	34.8	1.6	13542	18	V74465
40	34.6	1.6	979	20	X61816
41	34.6	1.6	1083	20	X61815
42	34.6	1.6	1295	20	X31077
43	34.6	1.6	2487	20	X20298
44	34.6	1.6	1230025	20	X91990
45	34.4	1.6	1899	20	Z52906

ALIGNMENTS

Streptococcus pneu
DNA encoding Cbpa
DNA encoding chol1
Streptococcus pneu
Clone PC3 encoding
Streptococcus pneu
DNA encoding a Bac
Bacillus thuringie
Mycoplasma genital
Human 5' EST isola
Borrelia burgdorfe
Enterococcus faeca
B. burgdorferi ant
B. burgdorferi ant
S. pneumoniae PspC
Borrelia burgdorfe
CDNA encoding a hu
Human pancreatic a
Staphylococcus aur
Staphylococcus aur
Staphylococcus aur
Staphylococcus aur
S. aureus RABR cod
Human synaptotomal
Staphylococcus aur
Enterococcus faeca
Staphylococcus aur
B. burgdorferi ant
B. burgdorferi ant
Nucleotide sequenc
Borrelia burgdorfe
Nucleotide sequenc
Human prostate tum

RESULT	ID	Score	Query Match Length	ID	Description
294275	1	2128	100.0	2128	Human PHELIIX CDNA
294275	2	48	2.3	3399	Chicken leucocytos
294275	3	42.8	2.0	11309	Borrelia burgdorfe
294275	4	42.8	2.0	910715	Borrelia burgdorfe
294275	5	41	1.9	11427	Enterococcus faeca
294275	6	40.4	1.9	164976	Methanococcus jann
294275	7	40	1.9	3763	Sequence encoding
294275	8	40	1.9	3763	Merzoite apical-e
294275	9	38.8	1.8	3029	Borrelia burgdorfe
294275	10	38.8	1.8	7953	Staphylococcus aur
294275	11	38.8	1.8	580073	Mycoplasma genital
294275	12	37.8	1.8	6846	Streptococcus pneu

```

XX WPI: 2000-237872/20.
DR P-PSDB: Y79269.
XX
PT Testis specific Helix Loop Helix proteins expressed in cancers and
PT useful for the prevention, diagnosis and treatment of prostate, bladder
PT and ovarian tumors -
XX
PS Claim 5; Fig 2A-D; 62pp; English.
XX
CC This sequence is that of cDNA encoding human PHELIx (see Y79269),
CC a novel basic Helix Loop Helix protein thought to act as a
CC transcription factor. PHELIx normally exhibits a testis-specific
CC expression pattern but is up-regulated in prostate and other types
CC of cancer. The cDNA clone, termed GPRC12 (ATCC 98956), was
CC isolated from a normal testis cDNA library using a suppression
CC subtractive hybridization method. The gene maps to chromosome
CC 13p13.1-13.3. The invention provides diagnostic and therapeutic
CC methods useful in the management of various cancers which express
CC PHELIx, including prostate cancer, bladder cancer, ovarian cancer
CC and testicular cancer, including therapies aimed at inhibition the
CC transcription, translation, processing or function of PHELIx. The
CC expression pattern of PHELIx suggests that is an ideal target for a
CC cancer vaccine approach to prostate cancer.
XX
SQ Sequence 2128 BP; 675 A; 463 C; 464 G; 526 T; 0 other:

Query Match      100.0%; Score 2128; DB 21; Length 2128;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaccgggggggggttggttcacacggcctcgtgcgtactggtctcttggtggcccttaa 60
DB 1 gaccgggggggggttggttcacacggcctcgtgcgtactggtctcttggtggcccttaa 60
QY 61 tgccttggtcctaaagggtcgtgaagggaagacgaggtcctctggtacactatg 120
DB 61 tgccttggtcctaaagggtcgtgaagggaagacgaggtcctctggtacactatg 120
QY 121 aaggaagagagaactacactcgaaggtgtgagacccaacgcctcgaacagcagttg 180
DB 121 aaggaagagagaactacactcgaaggtgtgagacccaacgcctcgaacagcagttg 180
QY 181 ctggagaagagccggaagaagtagtgaatggaacccgttatggaacacactggaatt 240
DB 181 ctggagaagagccggaagaagtagtgaatggaacccgttatggaacacactggaatt 240
QY 241 aatgatgaactagaaaaagcaaatgtttatctcaaggagaagaagtgaaaaaatccatg 300
DB 241 aatgatgaactagaaaaagcaaatgtttatctcaaggagaagaagtgaaaaaatccatg 300
QY 301 aaactcttcagatgacatctcttctatctcgtgtctataagaagaatgccaagtgaatt 360
DB 301 aaactcttcagatgacatctcttctatctcgtgtctataagaagaatgccaagtgaatt 360
QY 361 aaacacattacttaaacgcgttagaagaagaagaagactcttgaaagtccaagttaata 420
DB 361 aaacacattacttaaacgcgttagaagaagaagaagactcttgaaagtccaagttaata 420
QY 421 ctatgacttaacttgacaagaagaatcaaaaggtctacagaagaatcaacaatgaaagcgcg 480
DB 421 ctatgacttaacttgacaagaagaatcaaaaggtctacagaagaatcaacaatgaaagcgcg 480
QY 481 tacatactagctgaatgtctcaggttctggtttacatacaagttcttaaaagcaaca 540
DB 481 tacatactagctgaatgtctcaggttctggtttacatacaagttcttaaaagcaaca 540
QY 541 ggtgatacaactgcttagatgacagaagaatctagtgaacgcgaacaaatgacactta 600
DB 541 ggtgatacaactgcttagatgacagaagaatctagtgaacgcgaacaaatgacactta 600
QY 601 ttagttgagatgtgcactgttggtcactggtcgtactgtactgtacagaacatttcaaac 660

```

```

DB 601 ttagttgagatgtgcactgttggtcactggtcgtactgtactgtacagaacatttcaaac 660
QY 661 atagcaagaagtcaccatccatccatcagtgacaagaagaagcagcagcctttgtagat 720
DB 661 atagcaagaagtcaccatccatccatcagtgacaagaagaagcagcagcctttgtagat 720
QY 721 tgcataccaacatggttctcttaaggttgctcttcacaaatgacggagagctgaa 780
DB 721 tgcataccaacatggttctcttaaggttgctcttcacaaatgacggagagctgaa 780
QY 781 gccataagtttaattagatttgcgaaaaaaatacacatcactcgtttgtttata 840
DB 781 gccataagtttaattagatttgcgaaaaaaatacacatcactcgtttgtttata 840
QY 841 atccctgaaatttaaaaggttattctcagggacatgaaatgataatgcttaactgaa 900
DB 841 atccctgaaatttaaaaggttattctcagggacatgaaatgataatgcttaactgaa 900
QY 901 ccactgacaatggaaaaaataagtaagtgttaataatacttgacaacatgcccctcaaac 960
DB 901 ccactgacaatggaaaaaataagtaagtgttaataatacttgacaacatgcccctcaaac 960
QY 961 actgttaagactgaaaaacgcgaactgggctgaaagaacttgatggccctgcagaagttcc 1020
DB 961 actgttaagactgaaaaacgcgaactgggctgaaagaacttgatggccctgcagaagttcc 1020
QY 1021 taacagcaaacacccgtgatatattccctacagatctatttgcctccttgaaatttaag 1080
DB 1021 taacagcaaacacccgtgatatattccctacagatctatttgcctccttgaaatttaag 1080
QY 1081 aatggcaatgggttgatattaaatgcttcgttcgaagttcgaagaaacaaataatc 1140
DB 1081 aatggcaatgggttgatattaaatgcttcgttcgaagttcgaagaaacaaataatc 1140
QY 1141 tctctcttcataaagcaagaagaaacaaataatcgaagaaacaaataatcgtgtgag 1200
DB 1141 tctctcttcataaagcaagaagaaacaaataatcgaagaaacaaataatcgtgtgag 1200
QY 1201 cagctggtactctcttcgttcgttatgtaaaagggaagaagtgtggtcgttcgttct 1260
DB 1201 cagctggtactctcttcgttcgttatgtaaaagggaagaagtgtggtcgttcgttct 1260
QY 1261 gagcaacaagtgttatgtgttaataatataccggagagaagaatctccacgcgttaatg 1320
DB 1261 gagcaacaagtgttatgtgttaataatataccggagagaagaatctccacgcgttaatg 1320
QY 1321 cagattacaagaagcactcagaagcaacatgaggtttgtgaagaacaacaacaccatc 1380
DB 1321 cagattacaagaagcactcagaagcaacatgaggtttgtgaagaacaacaacaccatc 1380
QY 1381 gagctgtctctcccaaggcacctgtaatggaagcaggggaagaaacagtgtgtagagacttac 1440
DB 1381 gagctgtctctcccaaggcacctgtaatggaagcaggggaagaaacagtgtgtagagacttac 1440
QY 1441 tcccttgagagaagggctcccaatctcctactaataacgctgctggaatgggtctccactcct 1500
DB 1441 tcccttgagagaagggctcccaatctcctactaataacgctgctggaatgggtctccactcct 1500
QY 1501 gatgcagaagagctccttgatgaagcgtgtgagagttccatcaagctccgcctcagaagaat 1560
DB 1501 gatgcagaagagctccttgatgaagcgtgtgagagttccatcaagctccgcctcagaagaat 1560
QY 1561 gctatggtgatacattataaactcacatttccagtgacagcgtctctcgttaattctcttg 1620
DB 1561 gctatggtgatacattataaactcacatttccagtgacagcgtctctcgttaattctcttg 1620
QY 1621 catactgcaataataattcttaaaagtcaaccctcccaatgacatgctgtaacaat 1680
DB 1621 catactgcaataataattcttaaaagtcaaccctcccaatgacatgctgtaacaat 1680
QY 1681 cagaacatttcaatttactccttcagcagatggcccggtctcaagctctccctcggc 1740

```


	T05868 encodes a chicken leucocytozoan immunogenic protein, this DNA or a fragment of it can be used in a recombinant vaccine to immunise against chicken Leucocytosoan disease. The DNA is used in a vector CC and operatively linked to an expression regulatory sequence as in standard practice.
SQ	Sequence 3399 BP; 1577 A; 508 C; 798 G; 516 T; 0 other;
xx	
cc	
cc	
cc	
cc	
cc	
cc	
xx	
SQ	
	Query Match 2.3%; Score 48; DB 17; Length 3399; Best Local Similarity 48.2%; Pred.No. 0.0014; Matches 164; Conservative 0; Mismatches 175; Indels 1; Gaps 1.
Oy	168 gaaacagcagtgcctcggaagaagtcgccgcaagaatcacgtgatgaaggaaaaccgg-ttattgsg 226 Db 2227 gaaggaagaagaanaaaagrtacacatacyagaagaagaanaaagrtatatcatyaagaaaa 2286 Oy 227 acacaacttgtaataatgatgtacctagaaaaaaccaatgttactctaagtagaagaagtg 286 Db 2267 aaagaagaagtatatcatcayagaagaagaanaaagaagtacccaagaaagaagaaaaagaa 2346 Oy 287 gaaaaaaaaatcbtygaactcttcagatatgactatcttatcttgtctatygaaacgaaty 346 Db 2347 gaagcatacaactcyagaagaagaanaaagaagatatacacatcyagaagaagaanaaagaagta 2406 Oy 347 ccagtggaatctcttaaacaacttactttaacgcgtcagagaagaagaagaagctctltga 406 Db 2407 acatcgtgaagaagaanaaagaagaatlaaatccclyagaagaagaagaanaaagtacacaactga 2466

RESULT	3	
ID	X20250/c	
XX	X20250 standard; DNA; 111309 BP.	
XX		
AC	X20250;	
XX		
DT	04-MAY-1999 (first entry)	
XX		
DE	Borrelia burgdorferi polynucleotide sequence #3.	
XX		
KW	Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;	
KW	epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;	
KW	infection; diagnosis; characterisation; detection; ds.	
OS	Borrelia burgdorferi.	
XX		
FN	WO9858943-A1.	
XX		
PD	30-DEC-1998.	
XX		
PF	18-JUN-1998; 98WO-US12764.	
XX		
PR	03-SEP-1997; 97US-0057483.	
PR	20-JUN-1997; 97US-0050359.	
PR	22-JUL-1997; 97US-0053344.	
PR	22-JUL-1997; 97US-0053377.	
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
PA	(MEDI-) MEDIMUNE INC.	
PI	Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;	
PI	White OR;	
XX		
DR	WPI; 1999-081217/07.	

XX New isolated *Borrelia burgdorferi* nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of infections, particularly Lyme disease
 XX
 XX Claim 1: Page 738-800; 1128pp; English.
 XX
 CC X20248 to X20402 represent polynucleotide sequences isolated from
 CC *Borrelia burgdorferi* (Bb). Products derived from Bb can be used for
 CC the detection, diagnosis, characterisation, prevention and therapy of
 CC Bb infections, e.g. Lyme disease. They can also be used for the
 CC production of biosynthetic products, e.g. enzymes. *Borrelia* belongs
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.
 CC Spirochetes are pathogenic in humans and *Borrelia* causes epidemic and
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
 CC Lyme disease.
 CC
 XX Sequence 111309 BP; 35956 A; 13151 C; 19075 G; 43117 T; 10 other:
 SQ

Query Match 2.0%; Score 42.8; DB 20; Length 111309;
 Best Local Similarity 49.1%; Pred. No. 0.32; Indels 0; Gaps 0;
 Matches 113; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 205 tgaatggaacccgtatgagacacactgtaataatgatactagaagaagaat 264
 DB 75756 TTAATTAATAAATAGATATTAAATTAATTAATAGGCTTTTATGAAAAATTAAT 75697
 QY 265 tgttatctcaagagaagtggaanaaaatccatggaactcttcagatagactatctc 324
 DB 75696 TTTTAAATTAAGAGACTTAATAAATAATTTCTAACCATCTTGCACAAAAATTA 75637
 QY 325 tattcgctcctgacgaatgccagtgatcccttaaacacattacttaaacagctaga 384
 DB 75636 AATTTAATTAAGATTAAATTTATTTATTCGATTAAAAATCAAAAATTAACCTTTAA 75577
 QY 385 agaagaagaagaagactctggaagtcgaataactactgaacttaaac 434
 DB 75576 TAAAAAATTAATAGTTAATAAGATTAAGAGATATATATGAAATTAATAC 75527

RESULT 4
 X20248/c
 ID X20248 standard; DNA: 910715 BP.
 XX
 AC X20248;
 XX
 DT 04-MAY-1999 (first entry)
 XX
 DE *Borrelia burgdorferi* polynucleotide sequence #1.
 XX
 KW *Borrelia burgdorferi*; spirochete; bacterium; pathogen; Lyme disease;
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
 KW infection; diagnosis; characterisation; detection; ds.
 XX
 OS *Borrelia burgdorferi*.
 XX
 PN WO9858943-A1.
 XX
 PD 30-DEC-1998.
 XX
 PF 18-JUN-1998; 98WO-US12764.
 XX
 PR 03-SEP-1997; 97US-0057483.
 PR 20-JUN-1997; 97US-0050359.
 PR 22-JUL-1997; 97US-005344.
 PR 22-JUL-1997; 97US-0053377.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (MEDI-) MEDIMUNE INC.
 XX
 PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;
 PI White OR;

XX WPI: 1999-081217/07.
 XX
 XX New isolated *Borrelia burgdorferi* nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of infections, particularly Lyme disease
 XX
 XX Claim 1: Page 157-671; 1128pp; English.
 XX
 CC X20248 to X20402 represent polynucleotide sequences isolated from
 CC *Borrelia burgdorferi* (Bb). Products derived from Bb can be used for
 CC the detection, diagnosis, characterisation, prevention and therapy of
 CC Bb infections, e.g. Lyme disease. They can also be used for the
 CC production of biosynthetic products, e.g. enzymes. *Borrelia* belongs
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.
 CC Spirochetes are pathogenic in humans and *Borrelia* causes epidemic and
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
 CC Lyme disease.
 CC
 XX Sequence 910715 BP; 327171 A; 129646 C; 130753 G; 323091 T; 54 other:
 SQ

Query Match 2.0%; Score 42.8; DB 20; Length 910715;
 Best Local Similarity 49.1%; Pred. No. 1;
 Matches 113; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 205 tgaatggaacccgtatgagacacactgtaataatgatactagaagaagaat 264
 DB 665142 TTAATTAATAAATAGATATTAAATTAATTAATAGGCTTTTATGAAAAATTAAT 665083
 QY 265 tgttatctcaagagaagtggaanaaaatccatggaactcttcagatagactatctc 324
 DB 665082 TTTTAAATTAAGAGACTTAATAAATAATTTCTAACCATCTTGCACAAAAATTA 665023
 QY 325 tattcgctcctgacgaatgccagtgatcccttaaacacattacttaaacagctaga 384
 DB 665022 AATTTAATTAAGATTAAATTTATTTATTCGATTAAAAATCAAAAATTAACCTTTAA 664963
 QY 385 agaagaagaagaagactctggaagtcgaataactactgaacttaaac 434
 DB 664962 TAAAAAATTAATAGTTAATAAGATTAAGAGATATATATGAAATTAATAC 664913

RESULT 5
 X13102
 ID X13102 standard; DNA: 11427 BP.
 XX
 AC X13102;
 XX
 DT 19-MAR-1999 (first entry)
 XX
 DE *Enterococcus faecalis* genome contig SEQ ID NO:165.
 XX
 KW *Enterococcus faecalis*; contig; detection; Enterococcal infection;
 KW vaccine; attenuation; computer readable medium; ds.
 XX
 OS *Enterococcus faecalis*.
 XX
 PN WO9850555-A2.
 XX
 PD 12-NOV-1998.
 XX
 PF 04-MAY-1998; 98WO-US08965.
 XX
 PR 14-NOV-1997; 97US-0066009.
 PR 06-MAY-1997; 97US-0044031.
 PR 16-MAY-1997; 97US-0046655.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA Barash SC, Dillon PJ, Kunsch CA;
 XX
 PI WPI: 1999-045171/04.

PR	05-APR-1989;	89US-0334041.
XX		
PA	(UWNY-) NEW YORK UNIT.	
XX		
PI	Barwell JW, Galinski MR, Wertheimer SP;	
XX		
DR	WPI: 1990-334616/44.	
XX	P-PSDB: R07503.	
PT	Malarial apical end merozoite proteins and peptide(s) - used for	
XX	developing cpds. for treating, preventing and diagnosing malarial	
XX	infection	
PS	Claim 2; Fig 1a; 66pp; English.	
CC		
CC	A MAEP compound having a binding affinity for a Duffy blood group	
CC	antigen of primate red blood cells, is antigenic for the complete	
CC	protein, and may be used in diagnosis, treatment and vaccination	
CC	against invasion by P.vivax and P.knowlesi.	
XX		
SO	Sequence 3763 BP; 1796 A; 464 C; 611 G; 892 T; 0 other;	
	Query Match	1.9%; Score 40; DB 11; Length 3763;
	Best Local Similarity	49.4%; Pred. NO. 0.3;
	Matches 160; Conservative	0; Mismatches 160; Indels 4; Gaps 2;
QY	182 tggagaagtcgcgaagaagtgtgatafgaanaaccgttatgagacaaattgaat	241
DB	741 tgaacaacttaactgtgaattatgaataaaaagatgaattaaacttaattga	800
QY	242 aatgatacgaatgaanaagaattgtttatctcaaggagaagaatggaanaaa	301
DB	801 aataaagaatataagaacaaatgcacacgcgaatcagtaattcaaaagsga	860
QY	302 aactcttcgaatagactatctctattcgtctatgaacgagatgcagtgaa	361
DB	861 taanaattggtc---cttggaaaaatttaagcctaatgaggaacaaatcga	917
QY	362 aacacattacttaacacagctagaagaagaagaagacacttgaagaagtca	421
DB	918 ta-aacttaatgaataataatgaataataataagaacttcgaacaaactta	976
QY	422 tatgcacttaaaccttggaacaagaatcaaggcttaccagaatacaacatg	481
DB	977 aagacgcgcaaaaacaagaagtaglacaaaagtgaactatcataaacatg	1036
QY	482 acatacctagctgaatgtctcag	505
DB	1037 tcagtaataatttcagaaggaactgt	1060
RESULT	8	
ID	T80072	
XX	T80072 standard; DNA; 3763 BP.	
AC	T80072;	
XX		
DT	10-NOV-1997 (first entry)	
XX		
DE	Merozoite apical-end protein coding sequence clone 5.3.	
XX		
KW	Merozoite apical-end protein; MAEP; Plasmodium vivax; antigen; malaria;	
KW	human; erythrocyte; antimalarial vaccine; antibody; blood stage parasite;	
KW	Duffy blood group antigen; red blood cell; therapy; ds.	
XX		
OS	Plasmodium vivax.	
XX		
PN	US5646247-A.	
XX		
PD	08-JUL-1997.	
XX		
PF	05-APR-1989; 89US-0334041.	

XX	04-OCT-1991;	91US-0792865.
PR	05-APR-1989;	89US-0334041.
PR	06-APR-1989;	89US-0334270.
PR	03-APR-1990;	90MO-US01849.
XX	02-NOV-1990;	90US-0608639.
PR		
PA	(UNYK) UNIV NEW YORK STATE.	
PI	Barnwell JW, Galinski MR;	
DR	WPI; 1997-362995/73.	
XX	P-PSDB; W24575.	
PT	Plasmodium merozoite apical end protein - useful as antigen for	
PT	production of anti-malarial vaccines	
XX		
PS	Claim 1; Column 29-38; 68pp; English.	
CC	T80072 and T80073 represent the coding sequences for merozoite apical end	
CC	proteins (MAEP) isolated from two different Plasmodium vivax strains. The	
CC	encoded proteins are the antigens of the invention, and immunoreact with	
CC	antibodies against a native MAEP sequence. P. vivax is one of the four	
CC	malarial species that infects humans, and is difficult to target for a	
CC	vaccine, as it cannot be cultured in vitro. The preinvasion orientation	
CC	of malarial merozoites indicates that the apical end plays an important	
CC	role in the invasion process. The MAEP protein binds to the surface of	
CC	susceptible erythrocytes from P. vivax susceptible humans and primates,	
CC	and also binds to rabbit erythrocytes. The antigen can be used for the	
CC	production of anti-malarial vaccines. The antigens are involved in the	
CC	invasion process, and are immunochemically reactive with antibodies	
CC	raised against malaria (particularly P. vivax) blood stage parasites.	
CC	Synthetic proteins, polypeptides, peptide fragments and analogues of	
CC	these antigens can be used similarly. As the antigens specifically bind	
CC	to a Duffy blood group antigen (the antigen present on the surface of	
CC	susceptible mammalian red blood cells), and are necessary in the process	
CC	of invasion of red blood cells by merozoites, they can be used to inhibit	
CC	the invasion of red blood cells by a malarial organism. The antigens can	
CC	also be used in a method for inhibiting invasion of susceptible mammalian	
CC	blood cells by malarial merozoites, and in a method for inhibiting the	
CC	propagation of a malarial organism in susceptible red blood cells.	
XX		
SO	Sequence 3763 BP; 1796 A; 464 C; 611 G; 892 T; 0 other;	
Query Match	1.9%; Score 40; DB 18; Length 3763;	
Best Local Similarity	49.4%; Pred. No. 0.3;	
Matches 160; Conservative	0; Mismatches 160; Indels 4; Gaps 2	
QY	182 tggagaagatccgcaagaagttagtgaatggaaaaccggtatgagcacacttgatta	241
DB	741 tgaacaaatlaactcgtgaattatgtataaaagaatgaattaaatccatttaaetga	800
QY	242 aatgaatgaactgaaaagaacaaattgtttctctcaagaagagaagtggaaaaaatccatgta	301
DB	801 actaaagaagatataaagacaacaaatgacacacccgaatccagtaattcaaaagaggaaaaga	860
QY	302 aaccttcagatagactatcttcatctcgtctgtctatgatgaacgaatgccagtggaaatcccta	361
DB	861 taaatttgagtt---cttggaaaattttaagcctaattgaggagaagaatccgaataagt	917
QY	362 aacacattactaaagaagttagaagagaagaagaacacttgaagatcaagyygaataac	421
DB	918 taaccttaatgaaaaaataatgaataatataagaattctgaacaactcttaaaagatatag	976
QY	422 tatgcacttaaacctggaacaaagaatcaaaagcgttaccagaagaatcacaatgaaagccgt	481
DB	977 aagcagcagaaaaaacaagctagtaacaaaagtagaactattccataaacatgaaacaacta	1036
QY	482 acatacctagctgaaatgctcag	505
DB	1037 tcaagtaataatttcaaggaatctg	1060

Query Match	1.8%	Score 38.8	DB 20	Length 3029
Best Local Similarity	47.9%	Pred. NO. 0.59		
Matches 112	Conservative 0	Mismatches 122	Indels 0	Gaps 0
237	aattaaatgatygaactagagaagaacaattgtttatctcaaggagaagatgagaaaatcc	296		
Db	3019	ATTGAATTTCTTAAGCCCAATTAACCAAAAAATATTGAAATTAACCAAAAAAGAAATAGAAATAC	2960	
Qy	297	atggaactcttcagatgacatctcttcacgcgtgctatgaagcaatgcagtggaat	356	
Db	2959	AATAAAGTATTGGTACATATTATTCTTGAAATTATTTGGAATATATAAGAAACCAAAAAAAT	2900	
Qy	357	ccttaacacattacttaacacagctagaagaagaagaagactctggaagtcagtgta	416	
Db	2899	TCTTTAACCACAGGAAGAAATTAAACATAAGGCTATTTGAGATTTAAATATGGAAGTCC	2840	
Qy	417	aatactatgcactaaactagacaagaagaaagcttaccagaagatcaaca	470	

Db	2839	AATGACCCATATTGAACATATAAAAAAGGCAAAAGTAAAGTGAATGTCAAA	2786
RESULT	10		
ID	V74463	standard; DNA; 7953 BP.	
XX	V74463;		
AC	V74463;		
XX	16-MAR-1999	(first entry)	
DT	16-MAR-1999	(first entry)	
DE	Staphylococcus aureus contig SEQ ID #152.		
XX	Staphylococcus aureus.		
XX	Staphylococcus aureus.		
OS	Staphylococcus aureus.		
XX	Key	Location/Qualifiers	
FT	misc_feature	1081..1140	
FT	misc_feature	/tag= a	
FT	misc_feature	/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc_feature	2881..2940	
FT	misc_feature	/tag= b	
FT	misc_feature	/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc_feature	4681..4740	
FT	misc_feature	/tag= c	
FT	misc_feature	/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc_feature	6481..6540	
FT	misc_feature	/tag= d	
FT	misc_feature	/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
XX	EP786519-A2.		
PN	EP786519-A2.		
XX	30-JUL-1997.		
PD	30-JUL-1997.		
XX	07-JAN-1997;	97EP-0100117.	
PF	07-JAN-1997;	97EP-0100117.	
XX	05-JAN-1996;	96US-0009861.	
PR	05-JAN-1996;	96US-0009861.	
XX	(HUMA-) HUMAN GENOME SCI INC.		
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX	Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;		
PI	Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;		
XX	Rosen CA;		
PI	Rosen CA;		
XX	WPI; 1997-374922/35.		
DR	WPI; 1997-374922/35.		
XX	Polynucleotide(s) and proteins derived from Staphylococcus aureus		
PT	Polynucleotide(s) and proteins derived from Staphylococcus aureus		
XX	anti-S.aureus vaccines		
PT	anti-S.aureus vaccines		
XX	Claim 1; Page 770-774; 3271pp; English.		
PS	Claim 1; Page 770-774; 3271pp; English.		
XX	This sequence represents one of 5191 Staphylococcus aureus DNA sequences of the invention. The DNA sequences are recorded on a computer readable medium, preferably selected from a floppy or hard disk, random access memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using the S.aureus DNA sequences allows putative functions to be assigned so that protein-encoding or regulatory regions of commercial, therapeutic or		

FT	CDS	complement (46288..47422)
FT	FT	/*tag=
FT	FT	n
FT	FT	/label= MG039
FT	FT	/note= "Previously identified as MORF-19831 and
FT	FT	MORF-20106, the encoded protein shows 43.20
FT	FT	percentage identity to glycerol-3-phosphate
FT	FT	dehydrogenase (GUT2) from <i>S. cerevisiae</i> "
FT	FT	49377..49643
FT	FT	/*tag=
FT	FT	o
FT	FT	/label= MG041
FT	FT	/note= "The encoded protein shows 48.86 percentage
FT	FT	identity to phosphotransferase (ptsH) from <i>Mycobacteria</i>
FT	FT	<i>capricolum</i> "
FT	FT	50060..51520
FT	FT	/*tag=
FT	FT	p
FT	FT	/label= MG042
FT	FT	/note= "Previously identified as MORF-19832 and
FT	FT	MORF-20108, the encoded protein shows 41.92
FT	FT	percentage identity to spermidine/
FT	FT	putrescine transport ATP-binding protein
FT	FT	(potA) from <i>E. coli</i> "
FT	FT	51525..52382
FT	FT	/*tag=
FT	FT	q
FT	FT	/label= MG043
FT	FT	/note= "Previously identified as MORF-20110, the
FT	FT	encoded protein shows 26.51 percentage
FT	FT	identity to spermidine/putrescine transport
FT	FT	system permease protein (potB) from <i>E. coli</i> "
FT	FT	52366..53220
FT	FT	/*tag=
FT	FT	i
FT	FT	/label= MG044
FT	FT	/note= "Previously identified as MORF-20111, the
FT	FT	encoded protein shows 29.45 percentage
FT	FT	identity to spermidine/putrescine transport
FT	FT	system permease protein c (potC) from <i>E. coli</i> "
FT	FT	54658..55605
FT	FT	/*tag=
FT	FT	s
FT	FT	/label= MG046
FT	FT	/note= "Previously identified as MORF-20112, the
FT	FT	encoded protein shows 36.60 percentage
FT	FT	identity to sialoglycoprotease (9cp)
FT	FT	from <i>Pasteurella haemolytica</i> "
FT	FT	complement (56970..58310)
FT	FT	/*tag=
FT	FT	t
FT	FT	/label= MG048
FT	FT	/note= "Previously identified as MORF-19834,
FT	FT	MORF-20114 and MORF-20115, the encoded protein
FT	FT	shows 43.02 percentage identity to signal
FT	FT	recognition particle protein (ftn) from <i>B.</i>
FT	FT	<i>subtilis</i> "
FT	FT	58117..59079
FT	FT	/*tag=
FT	FT	u
FT	FT	/label= MG049
FT	FT	/note= "Previously identified as MORF-20114 and
FT	FT	MORF-20115, the encoded protein shows 44.78
FT	FT	percentage identity to purine-nucleoside
FT	FT	phosphorylase (deod) from <i>E. coli</i> "
FT	FT	59083..59754
FT	FT	/*tag=
FT	FT	v
FT	FT	/label= MG050
FT	FT	/note= "Previously identified as MORF-20117, the
FT	FT	encoded protein shows 83.03 percentage
FT	FT	identity to deoxyribose phosphate aldolase
FT	FT	(deoc) from <i>Mycobacteria pneumoniae</i> "
FT	FT	complement (64898..65731)
FT	FT	/*tag=
FT	FT	w
FT	FT	/label= MG056
FT	FT	/note= "Previously identified as MORF-20122, the
FT	FT	encoded protein shows 30.25 percent
FT	FT	identity to the protein disclosed in
FT	FT	GB:D26185..99 from <i>B. subtilis</i> "
FT	FT	complement (65713..66249)
FT	CDS	

[illegible]

XX	(HOMA-) HUMAN GENOME SCI INC.
PA	
PI	Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI	Kunsch CA, Rosen CA;
XX	
DR	WPI: 1998-272225/24.
XX	
PT	Computer-readable medium with recorded Streptococcus pneumoniae
PT	polynucleotide sequences - useful in diagnostic kits and assays; and
PI	pharmaceutical compositions and vaccines for Streptococcus
PT	pneumoniae
XX	
PS	Claim 1: Page 1169-1173; 1409pp; English.
XX	
CC	The present invention describes a computer readable medium which has
CC	the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC	on it, or a representative fragment or a sequence at least 95% identical
CC	to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC	to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC	pneumoniae. The present invention also describes an isolated nucleic acid
CC	molecule encoding a homolog of any of the fragments of the S.pneumoniae
CC	genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC	by a process comprising: (a) screening a genomic DNA library using as a
CC	probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC	to 391; identifying members of the library which contain sequences
CC	that hybridize to the target sequence and isolating the nucleic acid
CC	molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC	from an organism, amplifying nucleic acid molecules whose nucleotide
CC	sequence is homologous to amplification primers derived from the
CC	fragment of the S. pneumoniae genome to prime the amplification and
CC	isolating the amplified sequences. The computer readable medium can be
CC	used in a computer-based system for identifying fragments of the
CC	S. pneumoniae genome of commercial importance, or expression modulating
CC	fragments of the S. pneumoniae genome. Products from the present
CC	invention can be used in diagnosis kits and assays, and pharmaceutical
CC	compositions and vaccines for S. pneumoniae.
XX	
SO	Sequence 6846 BP; 2142 A; 1588 C; 1139 G; 1972 T; 5 other:
	Query Match 1.8%; Score 37.8; DB 19; Length 6846;
	Best Local Similarity 55.8%; Pred. No. 1.8;
Matches	72; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
OY	231 aaatcattggaagaacctcttagatagactatcttcatctgtctctatgaaagcaatgcag 350
DB	6004 ATACCACTTGAATTTAAACGATTCTTATTAACCTCATTAATAATCTTTAAGAAATAATAGA 5945
OY	351 tggaaatccctaaccatacttaacacagctgaagaagaagaagaagactcttgaaagtc 410
DB	5944 GGGTTTTGTCCAAGTATTAATTAGAGATTGAATAAGATATGAACAACCTCATTACAGA 5885
OY	411 aagtgaat 419
DB	5884 AAGTTAAAT 5876
.RESULT	13
ID	V27342
ID	V27342 standard; DNA; 1360 BP.
XX	
AC	V27342;
XX	
DT	02-OCT-1998 (first entry)
XX	
DE	Streptococcus pneumoniae SP0023 nucleotide.
XX	
KW	Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW	detection; pneumonia; otitis media; meningitis; ss.
XX	
OS	Streptococcus pneumoniae.
XX	
Key	Location/Qualifiers
PH	

FT	CDS		2..1360
FT		/tag= a	
FT		/product= "SP0023"	
FT		/note= "no stop codon given"	
PN		MO9818930-A2.	
PD		07-MAY-1998.	
PF		30-OCT-1997;	97WO-US19422.
PR		31-OCT-1996;	96US-0029960.
PA		(HUMA-) HUMAN GENOME SCI INC.	
PI		Choi GH, Hromockyj A, Johnson LS, Kunsch CA:	
DR		WPI: 1998-272224/24.	
P-PSDB:		W55081.	
PT		Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae - or their epitope-containing fragments, useful in protective or therapeutic vaccines, and for diagnosis	
PS		Claim 1; Page 55-56; 118pp; English.	
XX		The present sequence encodes a protein from Streptococcus pneumoniae. The nucleic acid sequence encoding the Streptococcus pneumoniae protein can be useful in vaccines for inducing protective antibodies against Streptococcus pneumoniae, for treatment or prevention of infection e.g. pneumonia, otitis media or meningitis. Probes based on the nucleic acid are used to detect Streptococcus infection (by usual hybridisation or amplification methods), also for isolating Streptococcus genes or their allelic variants. The protein can be used similarly to detect specific antibodies in standard immunoassays, especially for diagnosing or monitoring infections. Antibodies which bind the protein are used to detect corresponding antigens, to purify the protein and for passive immunisation (optionally coupled to a toxin). Vaccines are administered, e.g. by injection, orally or through the skin, typically at 0.01-1000 (especially 10-300) mu g/ml per dose.	
SQ		Sequence 1360 BP; 542 A; 264 C; 322 G; 232 T; 0 other:	
Query Match		1.7%; Score 37.2; DB 19; Length 1360;	
Best Local Similarity		47.8%; Pred. No. 1.1;	
Matches 108; Conservative		0; Mismatches 118; Indels 0; Gaps 0;	
OY	225	agaccacactgaatctaataatgatgaaactgagaagaacaattgttatctcaaggagaag	284
Db	537	agacgcagtcttaaaaagcagaagaagaagtcctaaacggaaagcacgaagaagataag	596
OY	285	tggaaaaaatccatgyaaactcttcagatlagactacttctattcgtytctatgaagaa	344
Db	597	ttaaagaaaacacagctgaacacacacacacgcgcgctccaagaagcagaanaaacag	656
OY	345	tggcgttggaatccttaaacacattacttaaacagctggaagaagaagaagactcttg	404
Db	657	cctccgtcccaaacacgaagaatccagcttgaacaaccaaagcagaanaaacacagctgac	716
OY	405	aaagtcaagtgaatatcatctacattcaacttgaacttggaaacaagaatacaaa	450
Db	717	aacaagctgaagaagacatctgctgtagatcagaagaagaataataa	762
RESULT	14		
ID	231402		
XX	231402	standard; DNA; 1969 BP.	
AC	231402:		
DT	07-FEB-2000	(first entry)	

Tue Jan 23 10:54:41 2001

us-09-389-000-1.rng

Page 12

Db 1263 ctccagctccaaaacgagagatccagctgcyaaacaacccaagcagaaaaaacccagctgac 1322

Qy 405 aaagtcagctgaataactatgcacttaaacctggaaacaagaatcaca 450

Db 1323 aaacaagctgaagaagactatgctccgtagatcagaagaagaatatga 1368

Search completed: January 19, 2001, 03:27:27
Job time: 15374 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2001, 22:07:46 ; Search time 1217.86 seconds
(without alignments)
12244.374 Million cell updates/sec

Title: US-09-389-000-1

Perfect score: 2128
Sequence: 1 gaccggggggcgttggggt.....ccacattatgaaaaaaa 2128

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_est40:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_esthum21:*
64: em_esthum22:*
65: em_esthum23:*
66: em_esthum24:*
67: em_esthum25:*
68: em_esthum26:*
69: em_esthum27:*
70: em_esthum28:*
71: em_esthum29:*
72: em_esthum30:*
73: em_esthum31:*
74: em_esthum32:*
75: em_esthum33:*
76: em_esthum34:*
77: em_esthum35:*
78: em_esthum36:*
79: em_esthum37:*
80: em_esthum38:*
81: em_esthum39:*
82: em_esthum40:*
83: em_esthum41:*
84: em_esthum42:*
85: em_esthum43:*
86: em_esthum44:*
87: em_esthum45:*
88: em_esthum46:*
89: em_esthum47:*
90: em_esthum48:*
91: em_esthum49:*
92: em_esthum50:*
93: em_esthum51:*
94: em_esthum52:*
95: em_esthum53:*
96: em_esthum54:*
97: em_esthum55:*
98: em_esthum56:*
99: em_esthum57:*
100: em_esthum58:*
101: em_esthum59:*
102: em_esthum60:*
103: em_esthum61:*
104: em_esthum62:*
105: em_esthum63:*
106: em_esthum64:*
107: em_esthum65:*
108: em_esthum66:*
109: em_esthum67:*
110: em_esthum68:*
111: em_esthum69:*
112: em_esthum70:*
113: em_esthum71:*
114: em_esthum72:*
115: em_esthum73:*
116: em_esthum74:*

117: em_estpl6:*
 118: em_estpl7:*
 119: em_estpl8:*
 120: em_estpl9:*
 121: em_estpl10:*
 122: em_estpl11:*
 123: em_estpl12:*
 124: em_estpl13:*
 125: em_estpl14:*
 126: em_estpl15:*
 127: em_estpl16:*
 128: em_estpl17:*
 129: em_estpl18:*
 130: em_estpl19:*
 131: em_estpl20:*
 132: em_estpl21:*
 133: em_estpl22:*
 134: em_estpl23:*
 135: em_estpl24:*
 136: em_estpl25:*
 137: em_estpl26:*
 138: em_estpl27:*
 139: em_estpl28:*
 140: em_estpl29:*
 141: em_estpl30:*
 142: em_estpl31:*
 143: em_estpl32:*
 144: em_estpl33:*
 145: em_estpl34:*
 146: em_estpl35:*
 147: em_estpl36:*
 148: em_estpl37:*
 149: em_estpl38:*
 150: em_estpl39:*
 151: em_estpl40:*
 152: em_estpl41:*
 153: em_estpl42:*
 154: em_estpl43:*
 155: em_estpl44:*
 156: em_estpl45:*
 157: em_estpl46:*
 158: em_estpl47:*
 159: em_estpl48:*
 160: em_estpl49:*
 161: em_estpl50:*
 162: em_estpl51:*
 163: em_estpl52:*
 164: em_estpl53:*
 165: em_estpl54:*
 166: em_estpl55:*
 167: em_estpl56:*
 168: em_estpl57:*
 169: em_estpl58:*
 170: em_estpl59:*
 171: em_estpl60:*
 172: em_estpl61:*
 173: em_estpl62:*
 174: em_estpl63:*
 175: em_estpl64:*
 176: em_estpl65:*
 177: em_estpl66:*
 178: em_estpl67:*
 179: em_estpl68:*
 180: em_estpl69:*
 181: em_estpl70:*
 182: em_estpl71:*
 183: em_estpl72:*
 184: em_estpl73:*
 185: em_estpl74:*
 186: em_estpl75:*
 187: em_estpl76:*
 188: em_estpl77:*
 189: em_estpl78:*

190: gb_gss25:*
 191: gb_gss26:*
 192: gb_gss27:*
 193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	612.2	28.8	617	97	AM961069 EST373036
2	545.6	25.6	624	97	AM961078 EST373045
3	484.6	22.8	491	4	AA293855 zt61d10.r
4	398.4	18.7	490	7	AA417643 zV04a08.r
5	373.2	17.5	422	143	R13043 yf71h06.r1
6	307.2	14.4	316	5	AA339260 EST44524
7	280.4	13.2	492	7	AA398348 zt61d10.s
8	244.2	11.3	503	111	BE65112
9	228.4	10.7	441	15	AI023533
10	180	8.5	500	146	W03238
11	168.4	7.9	394	13	AA861485
12	163.6	7.7	440	7	AA402245 zt65h06.r
13	143.4	6.7	360	15	AI028190
14	143.4	6.7	512	23	AI636904
15	124.6	5.9	282	32	AV205711
16	122.2	5.7	235	98	BB015390
17	122	5.7	170	183	B42447
18	116.6	5.5	274	30	AV044749
19	114.2	5.4	715	156	AO375951
20	113.4	5.3	226	98	BB015549
21	95.6	4.3	452	150	AQ196491
22	90.8	4.3	354	184	B92720
23	58.4	2.7	121	15	AI047238
24	50	2.3	878	190	CNS01878
25	50	2.3	1101	190	CNS00396
26	49.8	2.3	647	38	AV674451
27	49	2.3	1080	190	CNS01180
28	48.8	2.3	438	151	AQ267168
29	48	2.3	396	177	AZ277553
30	48	2.3	421	192	CNS03397
31	48	2.3	1101	190	CNS00182P
32	46.2	2.2	1101	190	CNS0118D
33	44.6	2.1	796	190	CNS0118D
34	44.2	2.1	1001	190	CNS0118D
35	44.2	2.1	1101	190	CNS0015H
36	43.8	2.1	1101	190	CNS0015H
37	43.6	2.0	559	156	AO371843
38	43.6	2.0	581	191	CNS0118D
39	43.6	2.0	768	191	CNS0118D
40	43	2.0	799	190	CNS0118D
41	43	2.0	1101	190	CNS003DO
42	43	2.0	1101	190	CNS006PA
43	43	2.0	1101	190	CNS01219
44	43	2.0	1309	108	BE420736
45	42.8	2.0	894	190	CNS018BC

ALIGNMENTS

RESULT 1
 AM961069
 LOCUS EST373036 617 bp mRNA
 DEFINITION EST373036 MAGF resequences, MAGF Homo sapiens cDNA, mRNA sequence.
 ACCESSION AM961069
 VERSION AM961069.1 GI:8150648
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 617)
Hegde, P., Qi, R., Abernathy, K., Dharp, S., Gaspard, R., Gay, C., Holt,
I. E., Seed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and
Quackenbush, J.

TITLE	JOURNAL
Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray	unpublished (2000)

The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel.: 301 838 3538

Location/Qualifiers

```

    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="MAGE resequences, MAGE
    /note="Vector: pB1uescriptSM"
BASE COUNT      211 a      116 c      127 g      163 t

```

BASE COUNT	211 a	116 c	127 g	163 t
ORIGIN				

Query Match	28.8%	Score 612.2	DB 97	Length 617
Best Local Similarity	99.5%	Pred. No. 1.3e-153		
Matches 614	Conservative	0	Mismatches 3	Indels 0
				Gaps 0

OY	295	ccatggaacaccttccgaatgacatctctcatccgtctctcatgaacgaatgacgagtgga	354
Db	1	CCATGGAACACTCTTCAGATACACTGTCTTCTATTTCGTCTATTATACCAATGCCAGTGGGA	60
OY	355	atccctaaacacatctacttaaacacagctagaagaagaaagaagacctctgaagtcgaat	414
Db	61	ATCCTTAAMCACTTACTTAATTAACAGCTGGAAGAAGAAAGAGGACCTCTTGAATCTCAAGT	120
OY	415	gaattacatctgaccttaactcgtgaacaagaatctcaaggctctacagaagaatcacaatga	474
Db	121	GAATACATCTACCTTAATTAACGTGGAACAGAAATCAAGGCTTACCAAGAAAGATCAACAAATGA	180
OY	475	aagccgtacatccctagctgaagaatgctcaaggtctggtttacatcaaatgtcttcaaaag	534
Db	181	ACGGCGTCAATCACTAGCTGAAATGTCACAGGGTTCGTGTACATCAAGTTCTTAAAG	240
OY	535	gcacaaggtgatacaactgcctctagatgtaagaagaatctagtgaanaagcgaataatgac	594
Db	241	GCAACAGGTGGATCAACTGCTTAGGATCAAGAGATATATGTGAAGAACGCAAAAATATGAC	300
OY	595	atcctatagcttggaagatgctacactgttggtcctacctgtcgtatctgatacgaagaacttt	654
Db	301	ATCTATTATAGTTGGAGATGTCACGTGGGCTACCTGGGTGATACGTGCACAAACTATTTT	360
OY	655	gcacaacatagaagaatctaccatcacctacgtgacacgaagagggacagacgacctttg	714
Db	361	GCAAACTAATGACGAAGATCACCATCACTCATGATACACAGAAGAGGACGACGCCCTTTTG	420
OY	715	gatgatgtcatatccaacatggtctctcttgaagtgctctcttcaactaagtgtgcggaagag	774
Db	421	GATGATTCATATTCACACATGTTCTCTTGAAGTGCTTCTTCACTAAGTGCCGAGAGAG	480
OY	775	ctggaagcgcatacaagtttaattagatgttggaaaaaagaataatacacatctcgttgcgt	834
Db	481	CTTGAAGCCATCAAGTATATATTAATTTGCAAAAAAGAAAATTAACACATCTCTTTTGT	540
OY	835	ttataatcccttgaanaattttaaaggtgttatcttcagggcatgtgaatgatattgcttta	894
Db	541	TTTATTAATCCTGAAAATTTTAAAAAGTGTATTTCAGGCGATGGAATGAGATATGCTTTA	600
OY	895	actgaacacatgacaat 911	

Db 601 ACTGAACCACTGACAAT 617

RESULT	2
AM961078	AM961078 624 bp mRNA EST 01-JUN-2000
LOCUS	ESTJ37304 MAGE resequences, MAGE Homo sapiens CDNA, mRNA sequence
DEFINITION	AM961078
ACCESSION	AM961078
VERSION	AM961078.1 GI:8150657
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE

1 (phases I to 624)

Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C., Holtz, T.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and

Quackenbush, J.

TITLE	Assessment of gene expression patterns in a model of colon tumor
JOURNAL	metastasis using a 19,200 element cDNA microarray
COMMENT	Unpublished (2000)
	Contact: John Quackenbush

The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel.: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 152
Seq primer: Forward.

FEATURES	Location/Qualifiers
source	1. .624

BASE COUNT	ORIGIN
209 a	117 c 128 g 170 t

Query Match	25.6%	Score 545.6	DB 97	Length 624
Best Local Similarly	96.3%	Pred. No. 1e-135		
Matches 591	Conservative	0	Mismatches 19	Indels 4
				Gaps 3

OY	235	ccatggaacaccccttcagaataacacacatctcattccgtgtcattgaagaagaagccagcagga	35
Db	1	ccatggaacaccccttcagaataacacatctcattccgtgtcattgaagaagaagccagcagga	60
OY	335	atcccttaacacattactacttaaacagctagaaagaagaagaagacctctgaagtcaagt	414
Db	61	atcccttaaacacattactacttaaacagctagaaagaagaagaagaagacacctctgaagtcaagt	120
OY	415	gaatatactatgcaacttaaacctgtaacaagaatcaaaagcgtaccagaagaatcaacaatga	474
Db	121	gaatatactatgcaacttaaacctgtaacaagaatcaaaagcgtaccagaagaatcaacaatga	180
OY	475	acgcgcgtacataccttagctgtaaaatgctcgaaggtctcgtgttaatacaagtctttaaag	533
Db	181	acgcgcgtacataccttagctgtaaaatgctcgaaggtctcgtgttaatacaagtctttaaag	240
OY	535	gcaacaggtgtgataacatgccttagatgtgacaagaagaatctagtgtgaaaacgcgaaaaatgac	594
Db	241	gcaacaggtgtgataacatgccttagatgtgacaagaagaagaatctagtgtgaaaacgcgaaaaatgac	300
OY	595	attctatagtttgtagatggtcactgctgtggctcactcctgtgatactgtacagaacatactt	654
Db	301	attctatagtttgtagatggtcactgctgtggctcactcctgtgatactgtacagaacatactt	360
OY	655	gcaaacatagagaagaatgcacatcaacatcaagtgtgacagaagaaggcagcagcgcttttg	714
Db	361	gcaaacatagagaagaatgcacatcaacatcaagtgtgacagaagaaggcagcagcgcttttg	420
OY	715	gagtattgcatattccaacatggtctccttgaaggtgcctcttaccataagtgccgaagag	774

Pb 421 GATGATTCATATCAACATGTTCTCTTGAAGTGCCCTTTTCACTAAGTCCGAGAG 480
 QY 775 cggagagccatcaagtaattag-attggcaaaagaataacacatcactg-cttg 832
 Db 481 CTGGAGGCCATCAAGTAATTAAGATTGGCAAAAGAAAATTCACAAATTCCTGTTTG 540
 QY 833 tttttaaacccttgaaaatttaaggtgtatctta--gggcaltgaaatgttg 890
 Db 541 TTTTAAATCCCTGAAAATTTTAAAGTTTGTTCATCAAGGCCCTGGAATGATTTTGT 600
 QY 891 tttaactgaaccac 904
 Db 601 TTTAACTTGAAACC 614
 RESULT 3
 AA293855 491 bp mRNA EST 12-AUG-1997
 LOCUS z61d10.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:726835
 DEFINITION 5', mRNA sequence.
 ACCESSION AA293855
 VERSION AA293855.1 GI:1941833
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 491)
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S.,
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
 Schellenberg, K., Stepien, M., Tan, F., Theising, B., White, Y., Wyile,
 T., Waterston, R. and Wilson, R.
 WashU-Merck EST Project 1997
 Unpublished (1997)
 CONTACT: Wilson RK
 WASHINGTON UNIVERSITY School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 833 Std Error: 0.00
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 446.
 FEATURES
 source
 1..491
 /organism="Homo sapiens"
 /db_xref="GDB:5923745"
 /db_xref="taxon:9606"
 /clone="IMAGE:726835"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pTZ193D-Pac (Pharmacia) with a modified
 polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech laboratories
 , Inc., and primed with a Not I - oligo(dT) primer [5',
 TGTTCACATCTGAGATGGAGCGGCCGCCCAATTTTCTTTTCTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pTZ193 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 166 a 99 c 115 g 111 t.
 ORIGIN
 Query Match 22.8% Score 484.6; DB 4; Length 491;
 Best Job: 1 Similarity 99.2%; Pred. No. 2.4e-119;
 Matches 67; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 13 gttggggttcacgcctgctgactgctctctggtggtgccttaattctctgctc 72

Db 1 GTTGGGCTTCACCGCCTGCGCTACTGCTTCTTGGTGGCCCTTAATGCTTGTC 60
 QY 73 taagtgctgaggaggaaagacgcggaggctctctgctgcacataagaagaagaga 132
 Db 61 TAAGTGCTGCCGGAAGAACCGGGAGGCTCTGTGCTCGACACTATGAAGAGAGAGA 120
 QY 133 aactacaacttcgacgggtgtgagcaccaacgcctgaaacagcagttgtggaagatc 192
 Db 121 AACTACAACTTCGACGGTGTGAGCACCAACCGCTGAACACAGAGTGTGGAGAGATC 180
 QY 193 cgcgaagtagtgaatggaagaaacccgttatgagacacacttgaatlaatgatgaatc 252
 Db 181 CGCAAGAACTAGTATGATGAAAAACCGCTATGAGACACAACTGAATTAATGATGAAT 240
 QY 253 agaaagcaaatgtttatctcgaaggagaagtggaaaaatccatggaactcttcaga 312
 Db 241 AGAAAGCAAAATGTTTACTCTCAAGAGAGAAAGTGAAGAAATCCATGAAACTCTTCAGA 300
 QY 313 taagactatcttctatctgctgtctatgaacgaatgcagtggaatccttaacacattact 372
 Db 301 TAACTATCTTCTTATTCGTCTATGACGAATGCCAGTGGAACTCTTAACACATTAAT 360
 QY 373 taaacagctagaagaagaagaagactcttgaagtgcaagtgaatatactatgacattaa 432
 Db 361 TAAACAGCTAGAAAGAGAAAGAGAGACTCTTGAAGTCAAGTGAATTAATGACATTA 420
 QY 433 actggaacagaatcaaaagcttaccagaagaatcaacatgaagcgcgactactaac 492
 Db 421 ACTGGAAACAGAAATCAAGGCTTACCAAGAAATCAACAAATGAAGCGCGTACATCTAGC 480
 QY 493 tgaatgcttc 503
 Db 481 TGAATGCTTC 491
 RESULT 4
 AA417643 490 bp mRNA EST 02-MAR-1998
 LOCUS zv04a08.r1 Soares_NHMPu_s1 Homo sapiens cDNA clone IMAGE:752630
 DEFINITION 5', mRNA sequence.
 ACCESSION AA417643
 VERSION AA417643.1 GI:2079462
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 490)
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S.,
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
 Schellenberg, K., Stepien, M., Tan, F., Theising, B.,
 White, Y., Wyile, T., Waterston, R. and Wilson, R.
 WashU-Merck human EST Project
 Unpublished (1997)
 CONTACT: Wilson RK
 WASHINGTON UNIVERSITY School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 950 Std Error: 0.00
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 465.
 FEATURES
 source
 1..490
 /organism="Homo sapiens"
 /db_xref="GDB:5975465"
 /db_xref="taxon:9606"
 /clone="IMAGE:752630"
 /clone_lib="Soares_NHMPu_s1"

/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
 /lab_host="DH10B"
 /note="Organ: mixed (see below); Vector: pT73D-Pac (Pharmacia) with a modified polylinker site.1: Not I; Site.2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NDH, pregnant uterus 2NDH, and fetal heart 2NDH) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

BASE COUNT 158 a 114 c 97 g 121 t
 ORIGIN

Query Match 18.7%; Score 398.4; DB 7; Length 490;
 Best Local Similarity 99.8%; Pred. No. 3.5e-96;
 Matches 399; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1727 gcttcccccgcgacgtccttctgtgtggccacagcgtcctacaccccaactg 1786
 Db 1 gcttcccccgcgacgtccttctgtgtggccacagcgtcctacaccccaactg 60
 QY 1787 tctgcaacagtttggcgatataaagcacatgttgaactcaccctcaccact 1846
 Db 61 tctgcaacagtttggcgatataaagcacatgttgaactcaccctcaccact 120
 QY 1847 actggcgcgagtttgacatctaggaagaatggaacaagaatgttgaagaagctccac 1906
 Db 121 actggcgcgagtttgacatctaggaagaatggaacaagaatgttgaagaagctccac 180
 QY 1907 ccaaaagcccaataataaagaagttgcatgttggctctgtgtaataatgacccaagc 1966
 Db 181 ccaaaagcccaataataaagaagttgcatgttggctctgtgtaataatgacccaagc 240
 QY 1967 tctgcagatagaagaagcagcagcgaagaagctggccacacactgtcaccatctcat 2026
 Db 241 tctgcagatagaagaagcagcagcgaagaagctggccacacactgtcaccatctcat 300
 QY 2027 acacacttgatcccccgcagcagcagagatcacaagaacaatgacctcagtgactaca 2086
 Db 301 acacacttgatcccccgcagcagcagagatcacaagaacaatgacctcagtgactaca 360
 QY 2087 ctctcttttccaaaataattcccaattatgaaaaa 2126
 Db 361 ctctcttttccaaaataattcccaattatgaaaaa 400

RESULT 5
 LOCUS R13043 422 bp mRNA EST 12-APR-1995
 DEFINITION yf71106.r1 Soares infant brain 1N1B Homo sapiens cDNA clone
 IMAGE:27805 5', mRNA sequence.
 ACCESSION R13043
 VERSION R13043.1 GI:766119
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman,
 M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
 Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston,
 R., Williamson, A., Wohlmann, P. and Wilson, R.
 The WashU-Merck EST Project
 Unpublished (1995)
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

TITLE
 JOURNAL
 COMMENT

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu
 Insert Size: 1034
 High quality sequence stops: 343 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1034 Std Error: 0.00
 Seq primer: M13Rpi
 High quality sequence stop: 343.
 FEATURES
 source location/Qualifiers
 1..422
 /organism="Homo sapiens"
 /db_xref="GDB:400152"
 /db_xref="taxon:9606"
 /clone="IMAGE:27805"
 /clone_1ib="Soares infant brain 1N1B"
 /sex="Female"
 /dev_stage="73 days post natal"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: whole brain; Vector: Lambda BA; Site.1: Not I; Site.2: Hind III; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5',
 ACTGGACAAATTCGGCGCCGACGAGAAATTTTCTTTTCTTTT 3'];
 double-stranded cDNA was ligated to Hind III adaptors
 (Pharmacia), digested with Not I and directionally cloned
 into the Not I and Hind III sites of the Lambda BA vector.
 Library went through one round of normalization. Library
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 138 a 80 c 95 g 109 t
 ORIGIN

Query Match 17.5%; Score 373.2; DB 143; Length 422;
 Best Local Similarity 96.4%; Pred. No. 2e-89;
 Matches 403; Conservative 0; Mismatches 13; Indels 2; Gaps 2;

QY 891 tttaactgaaccactgacaaatggaaaaaataatgtaatgtaataaactgacaaatg 950
 Db 1 tttaactgaaccactgacaaatggaaaaaataatgtaatgtaataaactgacaaatg 60
 QY 951 tccctcaaacactgttaagactgaaagcgaactggcctggaagaacttgatgccct 1010
 Db 61 tccctcaaacactgttaagactgaaagcgaactggcctggaagaacttgatgccct 120
 QY 1011 gcaagagtccttaagcgaacacactgggataatttccactgactatcttcgctctga 1070
 Db 121 gcaagagtccttaagcgaacacactgggataatttccactgactatcttcgctctga 180
 QY 1071 atcttaaggaatggaatggcgttgatgaatgaatgcttgcgtgcagagtcgagaanaa 1130
 Db 181 atcttaaggaatggaatggcgttgatgaatgaatgcttgcgtgcagagtcgagaanaa 240
 QY 1131 caaaaagatctctcttcttcaatcaagcgaagaagaactaagaagggaagaatcaata 1190
 Db 241 caaaaagatctctcttcttcaatcaagcgaagaagaactaagaagggaagaatcaata 300
 QY 1191 ttgctgtgagcagctgctactcctctgctgctatgtaaaaggagagaagaa-gaatgatcg 1249
 Db 301 ttgctgtgagcagctgctactcctctgctgctatgtaaaaggagagaagaa-gaatgatcg 360
 QY 1250 cttcaggtcttga-ggcaacagttgattatgtaaatatataccggygagaatctctc 1306
 Db 361 cttcaggtcttga-ggcaacagttgattatgtaaatatataccggygagaatctctc 418

RESULT 6
 LOCUS AA339260 316 bp mRNA EST 21-APR-1997
 DEFINITION EST44524 Fetal brain I Homo sapiens cDNA 5' end, mRNA sequence.
 ACCESSION AA339260
 VERSION AA339260.1 GI:1991678
 KEYWORDS EST.

Source	Organism	Human
REFERENCE	Authors	Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fulton, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-ali, C., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fene, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Frithman, J.L., Geoghegan, N.S., Glodok, A., Gnehm, C.L., Hanna, M.C., Heidelberg, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Mammos, S.M., Merrick, J.M., Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.R., Collins, E.J., Dinke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, M.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wang, J., Xu, C., Yu, G., Yu, G., Ruben, S.M., Dillion, P.J., Fannone, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.
TITLE	Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence	
JOURNAL	Nature 377 (6547 Suppl), 3-174 (1995)	
MEDLINE	96026280	
COMMENT	Contact: Kerlavage, AR Bioinformatics The Institute for Genomic Research 9712 Medical Center Drive, Rockville, MD 20850 USA Tel: 3018699056 Fax: 3018699423 Email: arkerlav@tigr.org	
FEATURES	Seq primer: M13 Reverse Location/Qualifiers 1..316 /organism="Homo sapiens" /db_xref="ATCC (Inhost):41055" /db_xref="taxon:9606" /clone_lib="Fetal brain I" /sex="female" /dev_stage="fetus, 24 wks" /note="Organ: brain; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"	
BASE COUNT	115 a 55 c 65 g 78 t 3 others	
ORIGIN		
Query Match	14.4% Score 307.2, DB 5; Length 316;	
Best Local Similarity	98.1%; Pred. No. 9,9e-72;	
Matches 309; Conservative	0; Mismatches 6; Indels 0; Gaps 0;	
263	attgcttatctcaaggaggaagtggaataacccatggaactcttcagatagactatct 322	
1	attttttttctcaagaggaagtggaataacccatggaactcttcagatagactgnt 60	
323	tctattcgtgtctcaagaggaagtggaactcagtggaatccttcaacaacattacttaacagagca 382	
61	tctattcgtgtctcaagaggaagtggaactcagtggaatccttcaacaacattactttaaagcgtca 120	
383	gaaggaagaagaagaagactcttgaagaagtcgaagtgaataactactgacacttaactgtaacaa 442	
121	gaaggaagaagaagaagactcttgaagaagtcgaagtgaataactactgacactttaaactgtaacaa 180	
443	gaatcaagaagcttaccagaagaatacaacaatgaagccggaataactactagctgaatgctct 502	
181	caatcaagaagcttaccagaagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgct 240	
503	cagggtctgtgttcaatcaagttctcaaaagcaacagggtgattcaactgcttagatg 562	

DB	241	CAGGCTTCGTTTACATCAAGTTTCAAAAGCCAGACAGGTGGATCAACTCCAGATG	300
QY	563	caaggaatctaatg 577	
DB	301	CAAGAGATCTAGTG 315	
RESULT	7		
LOCUS	AA398348/c		
DEFINITION	AA398348 492 bp mRNA	EST	12-AUG-1997
	z61d1d10.s1 Soares_testis_NHT Homo sapiens	cdna clone IMAGE:	726835
ACCESSION	AA398348		
VERSION	AA398348.1	GI:2051457	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1 (bases 1 to 492) Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucab, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, J., Schellander, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyll, T., Waterston, R. and Wilson, R.		
TITLE	WashU-Merck EST Project 1997		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810		
	Email: est@wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert length: 833 Std Error: 0.00 Seq primer: -41ml3 fwd. ET from Amersham High quality sequence stop: 372.		
FEATURES	Location/Qualifiers		
Source	1. 492		
	/organism="Homo sapiens"		
	/db_xref="GDB:5923745"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:726835"		
	/clone_idb="Soares_testis_NHT"		
	/sex="male"		
	/lab_host="DH10B"		
	/note="Vector: pT73d-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand CD was prepared from mRNA obtained from clonech laboratorci, Inc., and primed with a Not I - 0189(GGT) primer [5'].		
	TTGTACCAATCTGAAGTGGAGGCGGCCCAATTTTCTTTTCTTTT 3'] Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the NotI and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonalao. "		
BASE COUNT	126 a 91 c 92 g 183 t		
ORIGIN			
Query Match	13.2% Score 280.4; DB 7; Length 492;		
Best Local Similarity	94.8%; Pred. No. 1,8e-64;		
Matches 290; Conservative 0; Mismatches 16; Indels 0; Gaps			
QY	306	ctcagatagctatctctctattctgtytctatgaagatgcagtygaatccttaaca 365	
DB	492	CTTCAGATAGACTATCTTCTATTTCGTCTATGACCAATGCCAGTGGATCCTTAACA 433	
QY	366	cattacttaaacacatggaagaaagaaagacatcttgaagtcgaatgaatctatc 425	
DB	432	CATTACTTAAACACTGAGAACAAAGAGGAGCTCTTGAAAGTCAGTGAATFACTATG 373	
QY	426	cacttaaacctggaacgaatcaaggcttaccagaagatcaacaatgaaagccgtacat 485	

Db	Accession	Source	Organism	Reference Authors	Title	Journal	Comment
Db	61	ACTGCATTCCTGTGTGGGCGAGCTGTGCACAGCACACCCCACTGTCGTGGACAGT	120				
Qy	1801	ggcggtatataaagcacatgtttgaatccactctcaaccactactgagcgagttt	1860				
Db	121	GGGGGTATTAAATACCAATATTTGAAA--GCACCTTAGACCACCTAATTTGGGTGCGT	178				
Qy	1861	gacatctaggaaagagtggaacaaagaaatgatttgaagctccaccacaagactata	1920				
Db	179	GACCAATACAGGAAGAAATGGAACAAAGATGTCTTGAAGCTTACACCCAGAGACCTGATA	238				
Qy	1921	tcaaaagagttgagctgttctgtctctgtaaaatgcaactcaaaagcttctgcagata	1980				
Db	239	TCAAAAGGGTTTGATGTATGTATGGCTTCTGATAAATGTCTAAAGCTTCGACAGTAGAA	298				
Qy	1981	agaccagcagcgaaagaagctgagcacac-----actgtcaactatcttataaac	2030				
Db	299	AGACAGCAGCAAGAAAGTGTCTAGCTAGCGGTATGAGCTTTTATTAATAGTTATTAAT	358				
Qy	2031	actggtatcccgccagcagcagagctcaagaacaaatgagcctcagtgactacacc-	2089				
Db	359	ACTTGGATCTCC-CCACCAAGGTGCTCTAGAACCAATAGCCCTAGTGACTGTGACACT	417				
Qy	2090	tcttttctcaaaaatalttccacaatttatgaaaaa	2126				
Db	418	TTTTTCTCAAAACATATTCACAGTTATGACAAAA	454				
RESULT	9						
LOCUS	AI023533/c	441 bp	mRNA	EST	27-AUG-1998		
DEFINITION	ov79d11.s1 Soares_testis_NHT Homo sapiens						
LOCUS	AI023533						
VERSION	AI023533.1	GI:328577					
KEYWORDS	EST.						
SOURCE	human.						
ORGANISM	Homo sapiens						
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.						
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.						
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),						
COMMENT	Tumor Gene Index						
	Unpublished (1997)						
	Contact: Robert Strausberg, Ph.D.						
	Tel: (301) 496-1550						
	Email: Robert.Strausberg@nih.gov						
	CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldod						
	Ph.D.						
	CDNA Library Arrayed by: Greg Lennon, Ph.D.						
	DNA Sequencing by: Washington University Genome Sequencing Center						
	Clone distribution: NCI-CGAP clone distribution information can be						
	found through the I.M.A.G.E. Consortium/LLNL at:						
	www.bio.llnl.gov/bdrrp/image/image.html						
	Insert Length: 838 Std Error: 0.00						
	Seq Primer: -40m13 fwd. ET from Amersham						
	High quality sequence stop: 372.						
FEATURES	location/Qualifiers						
source	1. 441						
	/organism="Homo sapiens"						
	/db_xref="taxon:9606"						
	/clone="IMAGE:1643541"						
	/clone_1bp="Soares_testis_NHT"						
	/sex="male"						
	/lab_host="DH10B"						
	/note="Vector: pUT73D-Pac (Pharmacia) with a modified						
	polylinker: Site_1: Not I; Site_2: Eco RI; 1st strand cDNA						
	was prepared from mRNA obtained from Clontech Laboratories						
	, Inc., and primed with a Not I - oligo(dT) primer [5'						
	TGTACCAATCTGAAGGAGGAGCGGCCCAATTTTATTTTATTTT 3']						
	Double-stranded cDNA was ligated to Eco RI adaptors						
	(Pharmacia), digested with Not I and cloned into the Not I						
	and Eco RI sites of the modified pUT73 vector. Library						

BASE COUNT 85 a 87 c 84 g 138 t
 ORIGIN
 went through one round of normalization to Cct5, and was constructed by Bento Soares and M. Fatima Bonaldo. "

Query Match 7.9%; Score 168.4; DB 13; Length 394;
 Best Local Similarity 99.4%; Pred. No. 2.1e-34;
 Matches 169; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 415 gaattcttgccttaactggaagaataaagcgttcacagaagtcacatga 474
 |||||
 Db 394 GAAATCTACTGACCTTAAGTGAACAGATCAAGGCTTACAGAGATCAACATGA 335
 Qy 475 acgccctacatcctgctgaatgtctcagggtctgtgttcaatgaattcctaaag 534
 |||||
 Db 334 ACCCCCTACATCTACTGCTAAATGCTCTCAGGGTCTGTTTCAATCAATTTCTTAAAG 275
 Qy 535 gcacacagctgagatcaactgcctcctagatgcagaagaatctagtgaacgc 584
 |||||
 Db 274 GCACACAGTGTGATCACTGCTTGAATGCAAGATCTGTGAAAAACC 225

RESULT 12
 AA402245 440 bp mRNA EST 12-AUG-1997
 LOCUS 265h06.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:727259
 DEFINITION 5', mRNA sequence.
 ACCESSION AA402245
 VERSION AA402245.1 GI:2056883
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S.,
 Kucaba, T., Lacey, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
 Scheinberg, K., Stepien, M., Tan, F., Thaising, B., White, Y., Wyllie,
 T., Waterston, R. and Wilson, R.
 TITLE WashU-Merck EST Project 1997
 JOURNAL Unpublished (1997)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LNL: contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert length: 561 Std Error: 0.00
 Seq primer: -28m13 rev2 ET from Amersham.

FEATURES
 source
 1..440
 /organism="Homo sapiens"
 /db_xref="GDB:5924171"
 /db_xref="taxon:9606"
 /clone="IMAGE:727259"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech Laboratories
 , Inc., and primed with a Not I - oligo(dT) primer [5',
 TGTTCACATCTGAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization to Cct5, and was
 constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 94 a 110 c 135 g 101 t
 ORIGIN

Query Match 7.7%; Score 163.6; DB 7; Length 440;
 Best Local Similarity 97.6%; Pred. No. 4.2e-33;
 Matches 166; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 33 gccgtactggtcttgcctgaatgcttctgtctcctaagtgctgaggaaga 92
 |||||
 Db 9 GCTATCATGGCTTCTGGGTGGCCCTTAATGCTTGTGCTCTAAGGTCTGAGGGGAAGA 68
 Qy 93 cggggagagctctctggcctgacactatgaaggaagaactaacttgagctt 152
 |||||
 Db 69 CGCGGAGGCTCTTGGCTGACACTATGAAGGAAGAACTACAACTTGACCGGT 128
 Qy 153 gaaccacacccgctaaacacagctgtctggaagaagccgcaagaagt 202
 |||||
 Db 129 GAGCACCAACCGCTTAACACAGAGTGTGGAAGAAGTCCCAACAACT 178

RESULT 13
 AI028190 360 bp mRNA EST 27-AUG-1998
 LOCUS ov90c10.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1644594
 DEFINITION 3', mRNA sequence.
 ACCESSION AI028190
 VERSION AI028190.1 GI:3245499
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
 , Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www.bio.lnl.gov/bbrp/image/image.html
 Insert length: 874 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 320.

FEATURES
 source
 1..360
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1644594"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech Laboratories
 , Inc., and primed with a Not I - oligo(dT) primer [5',
 TGTTCACATCTGAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization to Cct5, and was
 constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 92 a 67 c 69 g 132 t
 ORIGIN

Query Match 7.0%; Score 148; DB 15; Length 360;
 Best Local Similarity 91.3%; Pred. No. 6.2e-29;

Matches 157; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 440 caagaatcaaggcttaccagaagaatcaacaataagcgcgtacatactgctgaatg 499
 |||||
 Db 360 CAAGAATCAAGGCTTACCAGAAGATCAACAATGACCCTACCTACCTGCTGAATG 301
 |||||

QY 500 tctcaaggttcgtttcaccatcaagttctcaaaaggcaacaggttgatcaactgcctatg 559
 |||||
 Db 300 TCTCAGGGTTGTGGTTTACATCAAGTTCTTAAAGGCAACAGGTGATCAACTGCCTTAGG 241
 |||||

QY 560 atgcaagaagaatctagtgaaaaacgaaaaatagacatcttattgttgaga 611
 |||||
 Db 240 ATGCAAGAAGATCTAGTGAAAACGTGAAGATATATATCCAGCTAAGCAGAGA 189
 |||||

RESULT 14
 A1636904 512 bp mRNA EST 26-APR-1999
 LOCUS ts90d06.x1 NCI_CGAP_G06 Homo sapiens cDNA clone IMAGE:2238539 3',
 DEFINITION mRNA sequence.
 ACCESSION A1636904
 VERSION A1636904.1 GI:4688234
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 512)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html
 Seq primer: -40UP from Gibco
 High quality sequence stop: 460.
 Location/Qualifiers
 1. 512
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2238539"
 /clone_1lb="NCI_CGAP_G06"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /note="Vector: p773D-PAC (Pharmacia) with a modified
 polynucleotide linker. Plasmid DNA from the normalized library
 NCI_CGAP_G04 was prepared, and ss circles were made in
 vitro. Following Hae purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from a pool of 5,000 clones made
 from the same library (clonoids 1257096-1258631,
 1469064-1470983, and 1475592-1476743). Subtraction by
 Bento Soares and M. Fatima Bonaldo.

BASE COUNT 113 a 111 c 111 g 177 t

ORIGIN

Query Match 6.7%; Score 143.4; DB 23; Length 512;
 Best Local Similarity 99.3%; Pred. No. 1.2e-27;
 Matches 144; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 442 agaatcaaggcttaccagaagaatcaacaataagcgcgtacatactgctgaatgctc 501
 |||||

Db 512 AGAATCAAGGCTTACCAGAAGATCAACAATGACCCTACCTACCTGAATGTC 453
 |||||

QY 502 tcaaggttcgtttcaccatcaagttctcaaaaggcaacaggttgatcaactgcctatg 561
 |||||

Db 452 TCAGGTTCTGGTTTACATCAAGTTCTTAAAGGCAACAGGTGATCAACTGCCTAGAT 393
 |||||

QY 562 gcaagaatctagtgaaaaacgcaa 586
 |||||

Db 392 GCAAGATCTAGTGAAAACGCA 368
 |||||

RESULT 15
 AV205711 282 bp mRNA EST 30-OCT-1999
 LOCUS AV205711
 DEFINITION AV205711 RIKEN full-length enriched, adult male testis Mus musculus
 cDNA clone 1700081I05 3', mRNA sequence.
 ACCESSION AV205711
 VERSION AV205711.1 GI:6146564
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 282)
 Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,
 Iishi, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai
 , C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M.,
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
 Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata
 , Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H.,
 Suzuki, H., Takahashi, F., Tateo, M., Tomioka, N., Tsunoda, Y.,
 Watanahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T.,
 Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.
 RIKEN Mouse ESTs (Konno, H., et al. 1999)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Yoshihide Hayashizaki
 Genome Exploration Research Group, Life Science Tsukuba Center,
 Genome Science Laboratory
 The Institute of Physical and Chemical Research (RIKEN), Genomic
 Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: +81-298-36-9013
 Fax: +81-298-36-9098
 Email: genome-res@ic.riken.go.jp,
 URL: http://genome.riken.go.jp/
 Sasaki, N., Izawa, M., Watanahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y.,
 Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y., and Hayashizaki
 , Y.
 Transcriptional sequencing: A method for DNA sequencing using RNA
 polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7): 3455-3460 (1998)
 Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki
 , Y., and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5): 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (http://genome.riken.go.jp) for
 further details.
 Location/Qualifiers
 1. 282
 /organism="Mus musculus"
 /strain="G57BL/6J"
 /db_xref="taxon:10090"
 /clone="1700081I05"
 /clone_1lb="RIKEN full-length enriched, adult male testis"
 /sex="male"
 /tissue_type="testis"
 /dev_stage="adult"
 /lab_host="SOLR"
 /note="Site_1: XhoI; Site_2: BamHI; cDNA library was

THIS PAGE BLANK (USPTO)